VISIT…

Vector Analysis for Computer Graphics

John Vince

Vector Analysis for Computer Graphics

Springer

John Vince, MTech, PhD, DSc
Emeritus Professor in Digital Media
Bournemouth University
UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2007923171

ISBN 978-1-84628-803-6 e-ISBN 978-1-84628-804-3

Printed on acid-free paper

© Springer-Verlag London Limited 2007

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

9 8 7 6 5 4 3 2 1

Springer Science+Business Media

springer.com

Dedication

This book is affectionately dedicated to my expanding family: Annie, Samantha, Anthony, Genny, Peter, Megan, and Monty.

Preface

In my last book, *Geometry for Computer Graphics*, I employed a mixture of algebra and vector analysis to prove many of the equations used in computer graphics. At the time, I did not make any distinction between the two methodologies, but slowly it dawned upon me that I had had to discover, for the first time, how to use vector analysis and associated strategies for solving geometric problems. I suppose that mathematicians are taught this as part of their formal mathematical training, but then, I am not a mathematician!

After some deliberation, I decided to write a book that would introduce the beginner to the world of vectors and their application to the geometric problems encountered in computer graphics. I accepted the fact that there would be some duplication of formulas between this and my last book; however, this time I would concentrate on explaining how problems are solved.

The book contains eleven chapters: The first chapter distinguishes between scalar and vector quantities, which is reasonably straightforward. The second chapter introduces vector representation, starting with Cartesian coordinates and concluding with the role of direction cosines in changes in axial systems. The third chapter explores how the line equation has a natural vector interpretation and how vector analysis is used to resolve a variety of line-related, geometric problems. Chapter 4 repeats Chapter 3 in the context of the plane.

At this point in the book, the reader has enough knowledge to tackle some standard problems encountered in ray tracing, such as reflections (Chapter 5) and intersections (Chapter 6).

Quaternions are the subject of Chapter 7, which is where I show how the clever combination of a scalar and a vector creates an object capable of rotating points about an arbitrary axis.

Chapter 8 introduces the idea of differentiating vector quantities, which are needed later on when we tackle bump mapping. Chapter 9 shows how vector analysis is used to describe projections, especially when the projection plane is oblique to the viewer's line of sight. Chapter 10 examines Gouraud and Phong shading as well as bump mapping, all of which utilize vectors as part of their algorithms. Finally, the book concludes with a short chapter on motion.

After reading this book, the reader should have a good understanding of how to employ vector analysis in solving a variety of geometric problems. I have found that the diagram used to summarize a problem's geometry often determines whether the solution drops out in a few lines or runs over several pages. For example, finding the formula for spherical interpolation involves the sine rule. And as you will discover in Chapter 7, the solution is extremely simple. However, if you start with the wrong diagram, the proof explodes into an algebraic nightmare!

I could not have written this book without the help of the books and papers cited in the reference section and the bibliography. And seeing how other authors approach problem solving has been very instructive, and I openly acknowledge their contribution.

I would like to thank my friends at Springer, Helen Desmond and Beverley Ford, for their regular emails (and lunches) asking when the manuscript would be ready. Without these reminders, the manuscript would still be sitting on my C drive! Thanks to them, it is in front of you, and I hope that you enjoy reading it and will refer to it from time to time.

John Vince
Ringwood

Contents

Preface ..... **vii**

1 Scalars and Vectors **1**

1.1 Introduction 1

1.2 Representing vector quantities 1

1.3 Non-collinear vectors 7

2 Vector Representation **11**

2.1 Introduction 11

2.2 Cartesian coordinates 11

2.3 Vector notation 13

2.4 Length of a vector 13

2.5 Vector algebra 14

2.6 Unit vectors 17

2.7 Rectangular unit vectors 18

2.8 Position vectors 19
2.8.1 Problem 1 19
2.8.2 Problem 2 20
2.8.3 Problem 3 21

2.9 Vector products 22
2.9.1 Scalar product 22
2.9.2 Vector product 27
2.9.3 Surface normals 33
2.9.4 The algebra of vector products 34

2.10 Triple products 36
2.10.1 Scalar triple product 36
2.10.2 The vector triple product 41

2.11 Perpendicular vectors 44
2.12 Interpolating vectors 48
2.12.1 Linear interpolation 48
2.12.2 Spherical interpolation 49
2.13 Direction cosines 52
2.14 Change of axial system 54
2.15 Summary 57

3 Straight Lines 61
3.1 Introduction 61
3.2 The parametric form of the line equation 61
3.3 The Cartesian form of the line equation 65
3.4 The general form of the line equation 68
3.5 2D space partitioning 69
3.6 A line perpendicular to a vector 74
3.7 The position and distance of a point on a line perpendicular to the origin . . . 77
3.7.1 The Cartesian form of the line equation 78
3.7.2 The parametric form of the line equation 80
3.8 The position and distance of the nearest point on a line to a point 81
3.8.1 The Cartesian form of the line equation 82
3.8.2 The parametric form of the line equation 84
3.9 The position of a point reflected in a line 86
3.9.1 The Cartesian form of the line equation 86
3.9.2 The parametric form of the line equation 88
3.10 A line perpendicular to a line through a point 89
3.10.1 The parametric form of the line equation 90
3.11 A line equidistant from two points 93
3.12 The equation of a line segment 94
3.13 The intersection of two straight lines 94
3.14 The point of intersection of two 2D line segments 97

4 The Plane 101
4.1 Introduction 101
4.2 The Cartesian form of the plane equation 101
4.3 The parametric form of the plane equation 103
4.4 A plane equation from three points 104

4.5 A plane perpendicular to a line and passing through a point 106
4.6 A plane through two points and parallel to a line 108
4.7 3D space partitioning 110
4.8 The angle between two planes 113
4.9 The angle between a line and a plane 114
4.10 The position and distance of the nearest point on a plane to a point 115
4.11 The reflection of a point in a plane 118
4.12 A plane between two points 120

5 Reflections 123
5.1 Introduction 123
5.2 A line reflecting off a line 123
5.3 A line reflecting off a plane 126

6 Intersections 129
6.1 Introduction 129
6.2 Two intersecting lines in $\mathbb{R}^2$ 130
6.3 A line intersecting a circle in $\mathbb{R}^2$ 133
6.4 A line intersecting an ellipse in $\mathbb{R}^2$ 138
6.5 The shortest distance between two skew lines in $\mathbb{R}^3$ 140
6.6 Two intersecting lines in $\mathbb{R}^3$ 142
6.7 A line intersecting a plane 144
6.8 A line intersecting a sphere 146
6.9 A line intersecting an ellipsoid 149
6.10 A line intersecting a cylinder 152
6.11 A line intersecting a cone 158
6.12 A line intersecting a triangle 160
6.13 A point inside a triangle 166
6.14 A sphere intersecting a plane 167
6.15 A sphere touching a triangle 172
6.16 Two intersecting planes 175

7 Rotating Vectors 179

7.1 Introduction 179

7.2 Rotating a vector about an arbitrary axis 179

7.3 Complex numbers 182
7.3.1 Complex number operations 183
7.3.2 Magnitude of a complex number 184
7.3.3 The complex conjugate 184
7.3.4 Graphical interpretation of complex numbers 185
7.3.5 i as a rotator 185
7.3.6 Unifying e, i, sin, and cos 186
7.3.7 Complex numbers as rotators 187

7.4 Quaternions 188

7.5 Quaternions as rotators 190

7.6 The complex conjugate of a quaternion 192

7.7 The norm of a quaternion 193

7.8 Inverse of a quaternion 195

7.9 Rotating vectors using quaternions 195

7.10 Representing a quaternion as a matrix 197

8 Vector Differentiation 201

8.1 Introduction 201

8.2 The derivative of a vector 201

8.3 The normal vector to a planar curve 204

8.4 The normal vector to a surface 206

9 Projections 213

9.1 Introduction 213

9.2 Perspective transform 213
9.2.1 Horizontally oblique projection plane 214
9.2.2 Vertically oblique projection plane 217
9.2.3 Arbitrary orientation of the projection plane 219

9.3 Pseudo fish-eye projection 223

10 Rendering 225

10.1 Introduction 225

10.2 Light sources 225

10.3 Local reflection models 227

10.4 Shading 230
10.5 Bump mapping 231

11 Motion 241
11.1 Introduction 241
11.2 Close encounters of the first kind 241
11.3 Close encounters of the second kind 243

Appendix A 247

Appendix B 249

References 253

Further Reading 255

Index 257

1 Scalars and Vectors

1.1 Introduction

Over thousands of years humans have developed various ways of counting things, whether they were oxen, days, trees, or stars. Various symbols, or strings of symbols, such as 9 or IX, have enabled us to count, record, compare, or communicate the number of things to our fellow human beings—such quantities are called *scalars*.

Initially, the natural numbers (1, 2, 3, 4 . . .) were sufficient to count most things, but eventually the class of *integers* emerged, which embraced the concepts of zero and negative numbers. Today we employ integers to count populations, real numbers to record the temperature, and transcendental numbers to express π, e, and trigonometric functions.

Scalar quantities embrace most things in our everyday lives such as height, age, temperature, page numbers, mass, distance, and tax rates. However, some things cannot be defined or measured by a single number. For instance, velocity or force requires a magnitude *and* a direction. So, too, does weight, which arises through one mass attracting another. In the case of objects on the earth, weight is determined by the object's mass and the earth's mass and is directed towards the earth's centre. The same object on the moon weighs less and is directed towards the moon's centre.

A natural step would be to employ two or more scalars to record such quantities, which is what happened about 150 years ago. After considerable debate, the term *vector* emerged to describe quantities possessing magnitude and direction, and a simple, yet powerful algebra was established to analyse and solve some challenging geometric problems.

1.2 Representing vector quantities

We can approach the subject of vectors from two directions: one employs an axiomatic basis; the second is intuitive. The axiomatic approach begins by defining the meaning of a vector space and the objects (vectors) that are members of this space, together with the axioms describing their manipulation. The intuitive approach is much more visual and uses *directed* or *oriented*

line segments to explain how vectors work. The second approach provides a rapid entry into the subject and will be sufficient for most readers. Appendix A provides a formal description of vectors for those readers who desire it.

Before vector notation was developed, problems involving forces were resolved using line segments, and forces were added together using the parallelogram rule shown in Fig. 1.1.

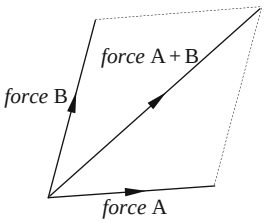


FIGURE 1.1.

The idea of using line segments to represent vectors has dominated the evolution of vector analysis, although it must be noted that some mathematicians have no need of such visual aids when they enter multi-dimensional worlds of abstract vector spaces! Fortunately for us, all of the problems considered in the following chapters are based upon simple two- or three-dimensional line segments.

A line segment is a perfect graphic for representing a vector, because its length represents magnitude, and its orientation represents direction. Figure 1.2 shows three identical vectors. They are identical because their length and orientation are equal, and their positions are irrelevant.

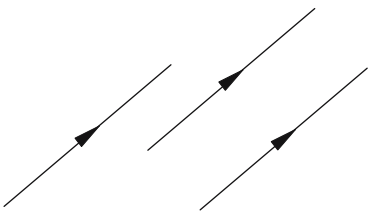


FIGURE 1.2.

The line segments in Fig. 1.2 have arrow heads, which define the vector's direction. Without the arrow the line segment could be pointing in either direction. Thus, a directed line segment is required to provide an unambiguous description of a vector.

Let us develop a notation for these line segments, one that will reflect the axioms that underpin vector algebra.

Figure 1.3(a) shows two points, A and B, connected by a line with an arrow pointing in the direction from A to B. This directed line segment is labeled $\overrightarrow{AB}$, where the arrow in the diagram confirms the direction from A to B, and the arrow on top of AB reminds us that it is a vector quantity.

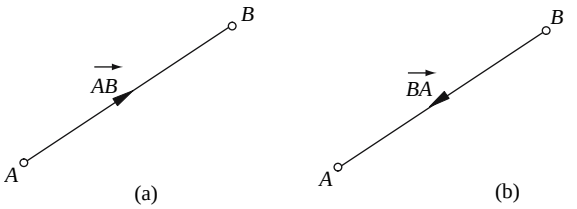


FIGURE 1.3.

If the line segment's arrow was pointing in the opposite direction, as shown in Fig. 1.3(b), it would be labeled $\overrightarrow{BA}$. What, then, is the difference between the two line segments $\overrightarrow{AB}$ and $\overrightarrow{BA}$? Well, the distance between the points is the same; just the direction has changed, and we can accommodate this difference by introducing a negative sign. For example,

$$\overrightarrow{BA} = -\overrightarrow{AB}$$

or conversely

$$\overrightarrow{AB} = -\overrightarrow{BA}.$$

The effect of the minus sign effectively switches the letters labeling the points.

If we adopt this notation, the following statements are valid descriptions of other line segments:

$$\overrightarrow{AC} = -\overrightarrow{CA}$$

and

$$-\overrightarrow{BC} = \overrightarrow{CB}.$$

The idea of following a path from one point to another is very useful when we begin solving problems with vectors. Along this path we identify the line segments (vectors) that locate some point, which eventually reveals the thing we are looking for.

Let's introduce a third point, C, and annotate the line segments as shown in Fig. 1.4.

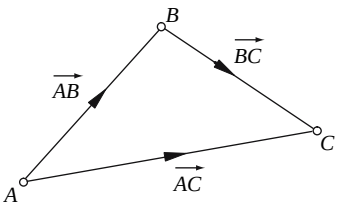


FIGURE 1.4.

We can move from A to C directly or indirectly. The direct path is $\overrightarrow{AC}$, and the indirect path is via B, i.e., $\overrightarrow{AB}+\overrightarrow{BC}$. Since both paths take us from A to C, we declare them equivalent, even though the Euclidean distances are not the same. Thus, we can state

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC} \tag{1.1}$$

or, going the opposite way,

$$-\overrightarrow{AC} = -\overrightarrow{AB} - \overrightarrow{BC}.$$

You may have noticed from Eq. (1.1) that the letters A and C in $\overrightarrow{AC}$ are the first and last letters in $\overrightarrow{AB}+\overrightarrow{BC}$. The Bs have effectively been cancelled. This pattern occurs frequently when we manipulate these labels.

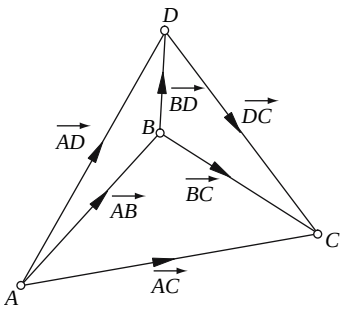


Figure 1.5.

Finally, let's add a fourth point, D, as shown in Fig. 1.5. The line segments connecting A, B, and C have not changed, but we have introduced line segments $\overrightarrow{AD}$, $\overrightarrow{BD}$, and $\overrightarrow{DC}$, which open up paths between any two points directly or indirectly. Although there may only be one direct route, there can be many indirect routes. For example, we can move from A to B in two ways:

A to C and then C to B

or

A to D and then D to B.

But wait! We can also go from A to B via C and *then* D. But hold on! We can also go from A to B via D and *then* C, i.e., A to D, D to C and then C to B.

They are all valid paths, and as more and more points are added, the number of indirect paths increases. But should we worry about all these indirect routes? Well, the answer is no. What is important is that we have a way of annotating a route, and this is influenced by the labels originally assigned to the line segments. For example, if we want to move from Q to P but the connecting line segment is labeled $\overrightarrow{PQ}$, then we use $-\overrightarrow{PQ}$ to represent $\overrightarrow{QP}$. When we

come to apply these techniques to real vectors, we will discover that some directions are more convenient than others.

Defining a line segment using $\overrightarrow{AB}$ or $-\overrightarrow{BA}$ is a useful notation we shall continue to employ. However, we can also give names to line segments, such as **a**, **n**, or **q**. Note that the boldface type differentiates a vector's name from the more familiar names given to scalar quantities such as x, y, or t. This is a widely accepted notation for referencing vectors and enables us to create diagrams such as the one shown in Fig. 1.6.

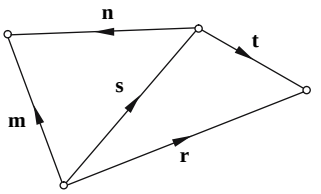


FIGURE 1.6.

From Fig. 1.6 we can make the following observations:

$$\mathbf{r} = \mathbf{s} + \mathbf{t}$$

$$\mathbf{s} = \mathbf{m} - \mathbf{n}$$

and

$$\mathbf{t} = \mathbf{n} - \mathbf{m} + \mathbf{r}.$$

One problem-solving strategy we use later on creates a chain of indirect vectors that eventually reveals the vector on the direct route. Such a situation is shown in Fig. 1.7, which permits us to write

$$\mathbf{p} = \mathbf{r} + \mathbf{s} + \mathbf{t} - \mathbf{n}.$$

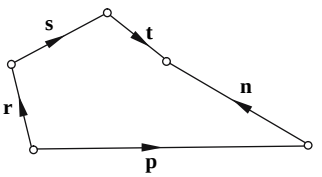


FIGURE 1.7.

You may be wondering why the vector **n** is reversed in Fig. 1.7. Why isn't it pointing in the same direction as **r**, **s**, and **t**? Well, when we come to labeling our vectors, certain directions are more convenient than others, which gives rise to such conflicts. However, it is nothing to worry about, because a simple minus sign resolves the problem.

The notation being developed here is very similar to that used in algebra. Indeed, vector algebra is virtually identical to ordinary algebra, apart from a couple of things that we will consider later on. For example, in algebra we can reason that if

$$x = 2 + b$$

and

$$b = c + d$$

then

$$x = 2 + c + d.$$

Similarly, if

$$x = -6$$

then

$$-x = 6.$$

We also know from algebra that if

$$x = 10 + b$$

then

$$10 = x - b.$$

So, can similar rules be used in vector algebra? Indeed, they can.

As we now know how to encode and interpret vector diagrams, let us see how we can identify points along a vector. Consider, then, the vector $\mathbf{r}$, as shown in Fig. 1.8. If vector $\mathbf{t}$ has an identical orientation but is half the length of $\mathbf{r}$, we can state

$$\mathbf{t} = \frac{1}{2}\mathbf{r}.$$

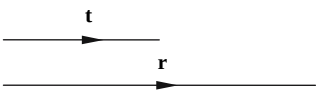


FIGURE 1.8.

But we could also reason that $\mathbf{r} = 2\mathbf{t}$, which is true and valid. And in general, we can state that

$$\mathbf{r} = \lambda\mathbf{p}$$

where λ is any scalar quantity. Therefore, we must get used to statements such as

$$\mathbf{p} = 2\mathbf{r} - 3\mathbf{s}.$$

These scalar multipliers provide a useful problem-solving strategy, as we shall see in later chapters.

When we add scalar quantities, we know that their sequence has no effect on the final result. For example, $2+6=8$ and $6+2=8$. Fortunately, this is the same for vectors. For instance, if $\mathbf{r}=\mathbf{s}+\mathbf{t}$, then we can also state that $\mathbf{r}=\mathbf{t}+\mathbf{s}$. This is shown graphically in Fig. 1.9(a), which illustrates the meaning of $\mathbf{r}=\mathbf{s}+\mathbf{t}$. But equally, we can reverse the vector sequence to create $\mathbf{r}=\mathbf{t}+\mathbf{s}$, as shown in Fig. 1.9(b).

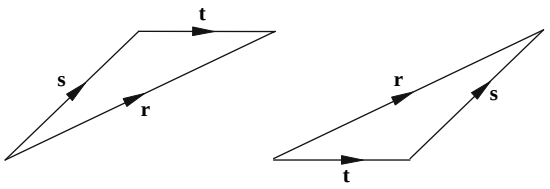


FIGURE 1.9.

When we subtract scalar quantities, their sequence is important. For example, $2-6=-4$, but $6-2=+4$. However, if we consider adding positive and negative scalar quantities together, we find that $2+(-6)=-4$ and $(-6)+2=-4$. This, too, has a vector equivalent, and it is worth investigating how this combination is represented graphically.

The operation $\mathbf{r}=\mathbf{s}+\mathbf{t}$ is shown in Fig. 1.9(a), but to draw $\mathbf{r}=\mathbf{s}-\mathbf{t}$, it is best to consider it as $\mathbf{r}=\mathbf{s}+(-\mathbf{t})$, as shown in Fig. 1.10. The process involves drawing the vector $\mathbf{t}$, reversing it to create $-\mathbf{t}$, and adding $\mathbf{s}$ to $-\mathbf{t}$.

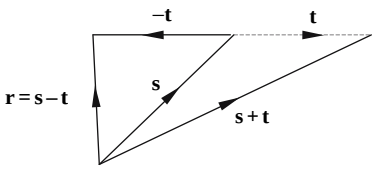


FIGURE 1.10.

1.3 Non-collinear vectors

When two vectors are *collinear*, we imply that the vectors possess the same orientation but could have different lengths. It means that one vector must be a scalar multiple of the other. If two vectors are *non-collinear*, then it is impossible for one to be a scalar multiple of the

other. For example, Fig. 1.11 shows triangle ΔABC constructed from three non-collinear vectors: $\mathbf{r} = \overrightarrow{AB}$, $\mathbf{s} = \overrightarrow{BC}$, and $\mathbf{r} + \mathbf{s} = \overrightarrow{AC}$. Superimposed is another triangle ADE with $a\mathbf{r} = \overrightarrow{AD}$, $b\mathbf{s} = \overrightarrow{DE}$, and $a\mathbf{r} + b\mathbf{s} = \overrightarrow{AE}$, where a and b are scalars. The vectors $a\mathbf{r}$ and $b\mathbf{s}$ are also non-collinear, because $\overrightarrow{AD}$ is parallel to $\overrightarrow{AB}$, and $\overrightarrow{DE}$ is parallel to $\overrightarrow{BC}$. This is written formally as $\overrightarrow{AD} \parallel \overrightarrow{AB}$ and $\overrightarrow{DE} \parallel \overrightarrow{BC}$, where the symbol $\parallel$ means *parallel to*.

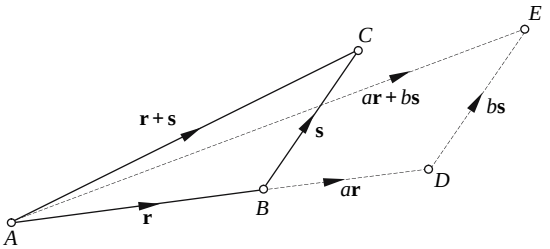


Figure 1.11.

Examine Fig. 1.11 very carefully and note that the route $\overrightarrow{AE}$ is unique. For no matter how we scale the vectors $\mathbf{r}$ and $\mathbf{s}$, and add them together, their sum will always produce a different resultant vector. If we want to move from A to E, only one pair of scalars exists to scale $\mathbf{r}$ and $\mathbf{s}$. This means that if ever we encounter a statement such as

$$a\mathbf{r} + b\mathbf{s} = c\mathbf{r} + d\mathbf{s}$$

where $\mathbf{r}$ and $\mathbf{s}$ are non-collinear, it can only mean one thing: $a = c$ and $b = d$. It is worth dwelling on this point until you really appreciate why this is so. This condition only applies to non-collinear vectors. However, such vector combinations are very common, and play an important role in problem solving.

We now know enough vector algebra to solve some simple geometric problems. So, consider proving that the opposite sides of a parallelogram are equal. We begin with the parallelogram $ABCD$ shown in Fig. 1.12, where $\mathbf{r} = \overrightarrow{AB}$ and $\mathbf{s} = \overrightarrow{AD}$. But as $AB \parallel DC$ and $AD \parallel BC$, we can conjecture that $\overrightarrow{DC} = u\mathbf{r}$ and $\overrightarrow{BC} = v\mathbf{s}$, where u and v are scalars.

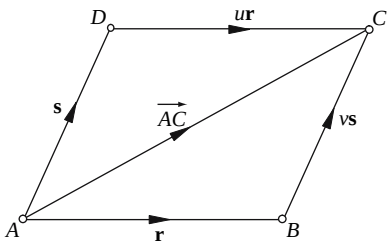


Figure 1.12.

We can now define the direct path $\overrightarrow{AC}$ in terms of two indirect paths:

$$\overrightarrow{AC} = \overrightarrow{AB} + \overrightarrow{BC}$$

and

$$\overrightarrow{AC} = \overrightarrow{AD} + \overrightarrow{DC}$$

therefore,

$$\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AD} + \overrightarrow{DC}.$$

Or, using vector names,

$$\mathbf{r} + v\mathbf{s} = \mathbf{s} + u\mathbf{r}.$$

Because these vectors are not collinear, we can equate coefficients and state that

$$1 = u \text{ and } v = 1.$$

If $u = 1$ and $v = 1$, $\overrightarrow{AB} = \overrightarrow{DC}$ and $\overrightarrow{AD} = \overrightarrow{BC}$, which means that their lengths are equal. Hence we have proved that the opposite sides of a parallelogram are equal.

Let's try another example and prove that the diagonals of a parallelogram bisect each other.

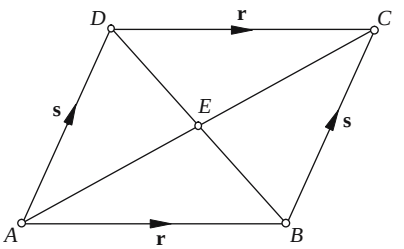


Figure 1.13.

Figure 1.13 shows a parallelogram $ABCD$ formed from the vectors $\mathbf{r}$ and $\mathbf{s}$, where $\mathbf{r} = \overrightarrow{AB} = \overrightarrow{DC}$ and $\mathbf{s} = \overrightarrow{AD} = \overrightarrow{BC}$. We can see that

$$\overrightarrow{AC} = \mathbf{r} + \mathbf{s} \tag{1.2}$$

and

$$\overrightarrow{DB} = \mathbf{r} - \mathbf{s}. \tag{1.3}$$

We observe that the diagonals intersect at E, which means that $\overrightarrow{AE}$ and $\overrightarrow{AC}$ are collinear. Therefore, we can state

$$\overrightarrow{AE} = \lambda\overrightarrow{AC} \tag{1.4}$$

where λ is some scalar.

Similarly, we can state that

$$\overrightarrow{DE} = \varepsilon \overrightarrow{DB} \tag{1.5}$$

where ε is some scalar.
We now want to prove that $\lambda = \frac{1}{2}$ and $\varepsilon = \frac{1}{2}$. We begin by rewriting (1.4) substituting (1.2):

$$\overrightarrow{AE} = \lambda(\mathbf{r} + \mathbf{s}) = \lambda\mathbf{r} + \lambda\mathbf{s}. \tag{1.6}$$

We can also rewrite (1.5) substituting (1.3):

$$\overrightarrow{DE} = \varepsilon(\mathbf{r} - \mathbf{s}) = \varepsilon\mathbf{r} - \varepsilon\mathbf{s}. \tag{1.7}$$

So, now we know $\overrightarrow{AE}$ and $\overrightarrow{DE}$. But we want a third equation that associates $\overrightarrow{AE}$ with $\overrightarrow{DE}$. Looking at Fig. 1.13, we see that in the triangle ΔAED we have

$$\overrightarrow{AE} = \overrightarrow{AD} + \overrightarrow{DE} = \mathbf{s} + \overrightarrow{DE}. \tag{1.8}$$

Therefore, substituting Eqs. (1.6) and (1.7) in Eq. (1.8), we obtain

$$\lambda\mathbf{r} + \lambda\mathbf{s} = \mathbf{s} + \varepsilon\mathbf{r} - \varepsilon\mathbf{s}$$

and

$$\lambda\mathbf{r} + \lambda\mathbf{s} = \varepsilon\mathbf{r} + (1 - \varepsilon)\mathbf{s}.$$

Now we have an equation linking λ, ε, $\mathbf{r}$, and $\mathbf{s}$. But as $\mathbf{r}$ and $\mathbf{s}$ are non-collinear, we can equate coefficients: therefore,

$$\lambda = \varepsilon \text{ and } \lambda = 1 - \varepsilon$$

which means that

$$\varepsilon = 1 - \varepsilon$$

and

$$\varepsilon = \frac{1}{2}$$

Also,

$$\lambda = \frac{1}{2}.$$

Therefore, the diagonals of a parallelogram bisect each other.

The time has now come to formalize these concepts and see how vectors can be encoded using Cartesian coordinates.

2 Vector Representation

2.1 Introduction

Vectors appeared in the mid-19th century, by which time analytic geometry had become established, especially the use of Cartesian coordinates. Consequently, it was a simple step to develop a numerical framework to represent vectors using Cartesian coordinates.

2.2 Cartesian coordinates

Figure 2.1 shows how this system works. The figure shows a vector with its *tail* located at (1, 1) and its *head* located at (3, 2). The words "head" and "tail" are often used to describe the two ends of a vector, with the head always associated with the direction of the vector's arrow. Cartesian coordinates provide a mechanism for representing a vector in terms of its horizontal and vertical Cartesian components. We can see that the horizontal component of the vector is

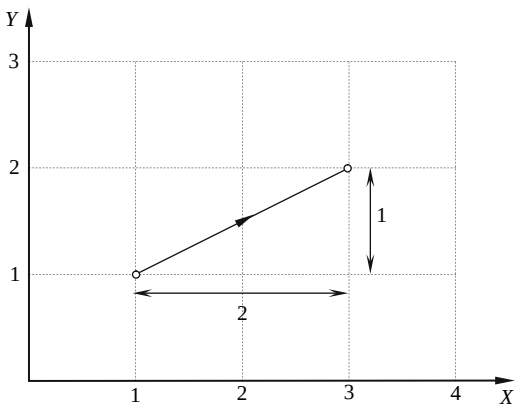


Figure 2.1.

obtained by subtracting the corresponding tail coordinates from the head coordinates: $3-1=2$, and its vertical component is obtained by subtracting $2-1=1$.

From the mid-19th to mid-20th centuries, two methods appeared to combine the x (horizontal) and y (vertical) components: one technique places the components as an ordered pair as $[x \quad y]$, and the other places them as $\begin{bmatrix} x \\ y \end{bmatrix}$. The former is called a *row vector*, while the latter is known as a *column vector*. In this text, we employ column vectors. However, when a column vector is referred to within a block of text, it is written as $\begin{bmatrix} x & y \end{bmatrix}^T$, which depicts a *transposed row vector*, i.e., a column vector, which saves space on the printed page.

No restrictions are placed upon vectors — they can be any length and point in any direction. However, as vectors are often computed from other vectors, it is possible to create a vector with no length — such a vector is called a *null* or *zero vector*. Figure 2.2 shows four vectors labeled **a**, **b**, **c**, and **d**, and Table 2.1 summarises their head and tail coordinates and x and y components.

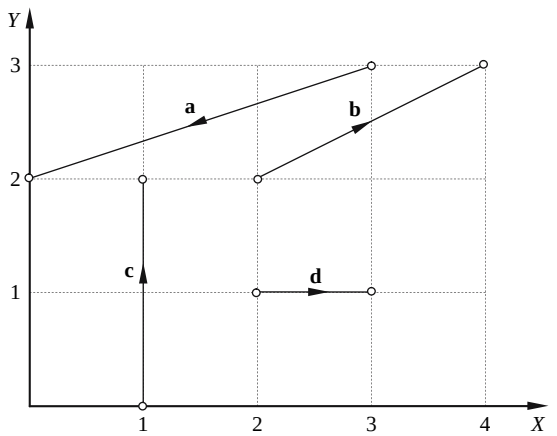


Figure 2.2.

Table 2.1

Vector	Head coordinates	Tail coordinates	x component	y component
a	(0, 2)	(3, 3)	−3	−1
b	(4, 3)	(2, 2)	2	1
c	(1, 2)	(1, 0)	0	2
d	(3, 1)	(2, 1)	1	0

Thus, the vectors are

$$\mathbf{a} = \begin{bmatrix} -3 \\ -1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \mathbf{d} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

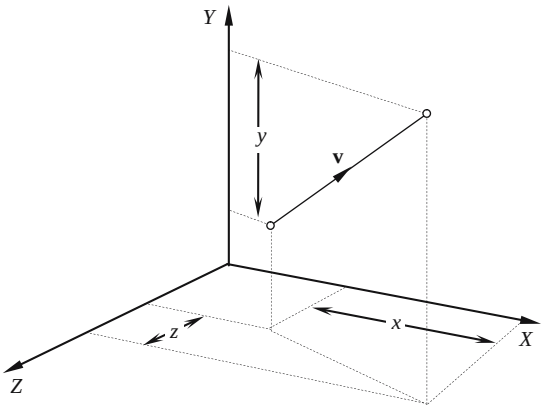


FIGURE 2.3.

So far our definition of a vector accommodates a 2D line segment but also embraces three-dimensional line segments. For example, Fig. 2.3 shows a 3D line segment with its Cartesian components, where the vector **v** is represented by

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Thus, a 2D vector has two components and a 3D vector has three components, which comprise an *ordered triple* of its Cartesian components.

A vector's orientation and length are determined completely by the sign and value of its Cartesian components. Fortunately, the length of any line segment is readily revealed by the theorem of Pythagoras; but we also require a way of expressing this symbolically.

2.3 Vector notation

Boldface type is used for vector names. This is a universal convention and helps distinguish scalars $(x, y, z, r, s, s, t \ldots)$ from vectors $(\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{n}, \mathbf{p}, \mathbf{q}, \ldots)$. However, when we are dealing with a line segment formed from two points, such as A and B, the associated vector is annotated as $\overrightarrow{AB}$ or **AB**, which represents a vector with its tail at A and head at B.

2.4 Length of a vector

We know that scalars can be positive or negative, and when we are only interested in the absolute value of a scalar s (i.e., sign independent), we use the notation $|s|$, which effectively strips away its sign. The same notation is used to represent the length of a vector, i.e., $|\mathbf{a}|$. It is also possible to use $\left|\overrightarrow{AB}\right|$ or $|\mathbf{AB}|$. Just to confuse matters, double vertical lines can be used

instead of single lines: $\left\| \overrightarrow{AB} \right\|$ or $\| \mathbf{AB} \|$, which is the convention used in this text to distinguish it from the scalar operation.

If a 2D vector $\mathbf{v}$ is given by

$$\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$$

its length is given by

$$\| \mathbf{v} \| = \sqrt{x^2 + y^2}.$$

Similarly, if a 3D vector is given by

$$\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

its length is given by

$$\| \mathbf{v} \| = \sqrt{x^2 + y^2 + z^2}.$$

These calculations are based upon the theorem of Pythagoras and are illustrated in Fig. 2.4.

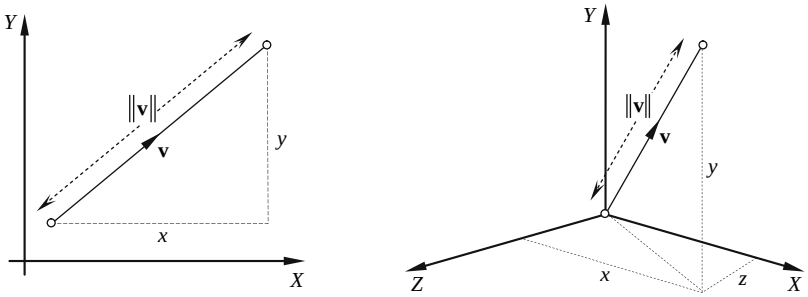


FIGURE 2.4.

The length of a vector is a scalar. Thus, $\| \mathbf{v} \|$ is a scalar and can be used alongside any other scalar and manipulated in a similar manner.

Let us now proceed in developing an algebra that shows how vectors are manipulated arithmetically and integrated with scalars. It is convenient to illustrate this in two dimensions, but the principles also apply to three dimensions.

2.5 Vector algebra

Reversing the sign of a scalar is achieved by multiplying it by -1. For example, $-1 \times (+3) = -3$ and $-1 \times (-3) = 3$. Similarly, a vector's direction is reversed by multiplying its components by -1. For example, if

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

then

$$-\mathbf{v} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}.$$

Figure 2.5 shows the graphic interpretation of this sign reversal.

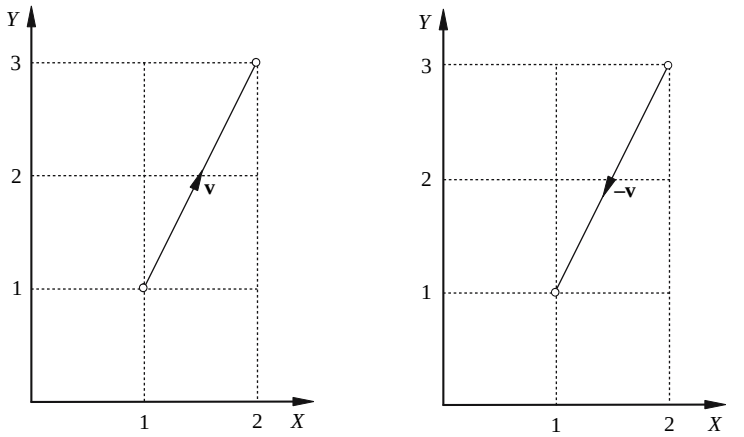


FIGURE 2.5.

A vector can be multiplied by any positive or negative scalar to make it larger or smaller. For example, we could double a vector by multiplying it by 2: given

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

then

$$2\mathbf{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

and

$$\tfrac{1}{2}\mathbf{v} = \begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}.$$

Equally, vector addition and subtraction obey the same rules as scalars. For example, given two vectors

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \text{ and } \mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

then

$$\mathbf{v} + \mathbf{w} = \begin{bmatrix} 1+2 \\ 2+1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

In contrast,

$$\mathbf{v} - \mathbf{w} = \begin{bmatrix} 1-2 \\ 2-1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

and

$$\mathbf{w}-\mathbf{v}=\begin{bmatrix}2-1\\1-2\end{bmatrix}=\begin{bmatrix}1\\-1\end{bmatrix}.$$

Vector addition and subtraction also have a simple graphical interpretation. In fact, vector addition uses the parallelogram rule shown in Fig. 1.1. Figure 2.6 illustrates how vectors **v** and **w** are added together. It is obvious that the components of the vector addition are obtained by adding the individual components together. Remember that these vectors do not have any absolute position and can be positioned anywhere in space.

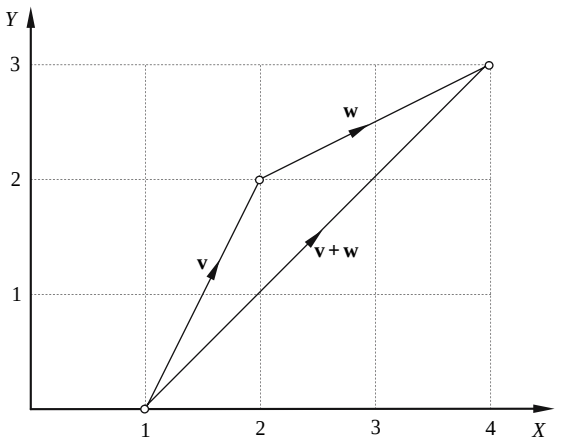


Figure 2.6.

Vector subtraction is illustrated in Fig. 2.7. To simplify the process, it is best to regard $\mathbf{v}-\mathbf{w}$ as $\mathbf{v}+(-\mathbf{w})$.

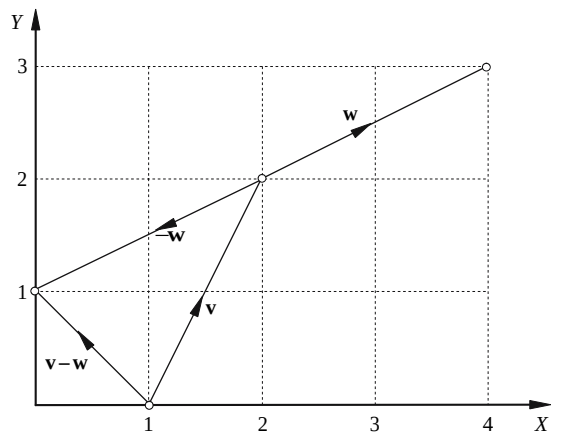


Figure 2.7.

The multiplication and division of two vectors are not so obvious. To begin with, dividing one vector by another does not seem to mean anything and has no definition, whereas vectors can be multiplied together in two ways, and we explore these products later.

The laws associated with scalars are simple and very familiar. Fortunately, the same laws apply to vectors; the only difference between the two systems is in the product of two vectors, which have still to be covered. Table 2.2 gives examples of the two systems.

Table 2.2

Law	Scalar algebra	Vector algebra
Commutative law for addition	$a+b=b+a$	$\mathbf{a}+\mathbf{b}=\mathbf{b}+\mathbf{a}$
Associative law for addition	$a+(b+c)=(a+b)+c$	$\mathbf{a}+(\mathbf{b}+\mathbf{c})=(\mathbf{a}+\mathbf{b})+\mathbf{c}$
Commutative law for multiplication	$ab=ba$	$a\mathbf{b}=\mathbf{b}a$
Associative law for multiplication	$a(bc)=(ab)c$	$a(b\mathbf{c})=(ab)\mathbf{c}$
Distributive law for multiplication	$(a+b)c=ac+bc$	$(a+b)\mathbf{c}=a\mathbf{c}+b\mathbf{c}$
	$a(b+c)=ab+ac$	$a(\mathbf{b}+\mathbf{c})=a\mathbf{b}+a\mathbf{c}$

2.6 Unit vectors

A *unit vector* has a length of 1. Unit vectors greatly simplify problem solving; therefore, we should understand how they are created.

For example, if $\mathbf{v}=\begin{bmatrix}x & y\end{bmatrix}^T$ and is a unit vector, we are implying that $\|\mathbf{v}\|=1$, i.e., $\sqrt{x^2+y^2}=1$. It is normally written as $\hat{\mathbf{n}}$, where the circumflex reminds us of the vector's unit length. We already know that a vector's length can be controlled by a scaling factor. Therefore, any vector has to be some multiple of a unit vector:

$$\mathbf{v}=\lambda\hat{\mathbf{v}}.$$

But, surely, $\lambda=\|\mathbf{v}\|$. Therefore, we can write

$$\mathbf{v}=\|\mathbf{v}\|\hat{\mathbf{v}}$$

which leads us to

$$\hat{\mathbf{v}}=\frac{\mathbf{v}}{\|\mathbf{v}\|}.$$

For example, given

$$\mathbf{v}=\begin{bmatrix}3\\4\end{bmatrix}$$

we can make it a unit vector by dividing its components by $\|\mathbf{v}\|$, which, in this case, is

$$\sqrt{3^2+4^2}=5.$$

i.e.,

$$\hat{\mathbf{v}}=\begin{bmatrix}0.6\\0.8\end{bmatrix}.$$

2.7 Rectangular unit vectors

A very powerful feature of vector algebra emerges from the previous definition of a vector, in that it is possible to express a vector as the sum of unit vectors aligned with the rectangular Cartesian axes. The following description moves to three dimensions but is equally applicable in two dimensions. Note also that throughout this text we will employ a right-handed system of 3D axes.

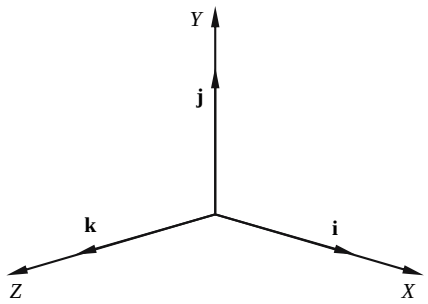


FIGURE 2.8.

We begin by defining three *rectangular unit vectors* **i**, **j**, and **k**, parallel with the x-, y-, and z-rectangular Cartesian axes, respectively, as shown in Fig. 2.8, where

$$\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad \mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad \mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Consequently, any vector $\mathbf{v} = \begin{bmatrix} a & b & c \end{bmatrix}^T$ can be written as

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

which provides us with a simple algebraic mechanism for manipulating vectors.

For example, we can reverse **v**:

$$-\mathbf{v} = -a\mathbf{i} - b\mathbf{j} - c\mathbf{k}.$$

We can double the length of **v**:

$$2\mathbf{v} = 2a\mathbf{i} + 2b\mathbf{j} + 2c\mathbf{k}.$$

And if we have two vectors,

$$\mathbf{v} = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}$$

and

$$\mathbf{w} = 5\mathbf{i} + 6\mathbf{j} + 7\mathbf{k}$$

we can add them together:

$$\mathbf{v} + \mathbf{w} = (2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}) + (5\mathbf{i} + 6\mathbf{j} + 7\mathbf{k}) = 7\mathbf{i} + 9\mathbf{j} + 11\mathbf{k}.$$

Soon we will discover how to multiply two vectors using this notation.

2.8 Position vectors

Imagine a point P in space with coordinates (x, y, z). Obviously, there exists a vector whose tail is at the origin and head is at P. Such a vector is called a *position vector* as it fixes the position of P. What is useful, though, is that the vector's rectangular components are identical to the Cartesian coordinates of the point, i.e., $\mathbf{p} = x\mathbf{i} + b\mathbf{j} + c\mathbf{k}$.

Before proceeding, let's take stock of what we have learned by applying this knowledge to solve some simple problems.

2.8.1 Problem 1

An object is subjected to two forces F_1 and F_2, where F_1 acts horizontally left to right, while F_2 acts vertically downwards, as shown in Fig. 2.9. The problem is to find the magnitude and direction of the resultant force and the force that would keep the object in equilibrium.

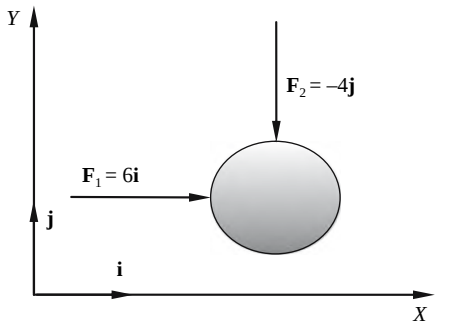


Figure 2.9.

We begin by defining the forces as two vectors:

$$\mathbf{F}_1 = 6\mathbf{i}$$

and

$$\mathbf{F}_2 = -4\mathbf{j}.$$

Therefore, the resultant force is

$$\mathbf{F}_1 + \mathbf{F}_2 = 6\mathbf{i} - 4\mathbf{j}.$$

The magnitude (length) is

$$\| \mathbf{F}_1 + \mathbf{F}_2 \| = \sqrt{6^2 + 4^2} = \sqrt{52} = 7.211.$$

The direction of the force can be specified relative to the horizontal x-axis as shown in Fig. 2.10.

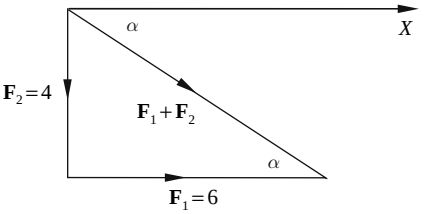


FIGURE 2.10.

From Fig. 2.10, we find that

$$\alpha = \tan^{-1}\left(-4/6\right) = -33.69^\circ \text{ or } 326.31^\circ.$$

For the object to be in equilibrium the total force must be zero, which means that a third force must be applied equal and opposite to $\mathbf{F}_1 + \mathbf{F}_2$, i.e.,

$$-(\mathbf{F}_1 + \mathbf{F}_2) = -(6\mathbf{i} - 4\mathbf{j}) = -6\mathbf{i} + 4\mathbf{j}.$$

Its direction is

$$\beta = 180^\circ + \tan^{-1}\left(4/-6\right) = 146.31^\circ.$$

2.8.2 Problem 2

Let's prove that the addition of two vectors is commutative, i.e., $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.

We begin by defining two vectors, **a** and **b**, as shown in Fig. 2.11.

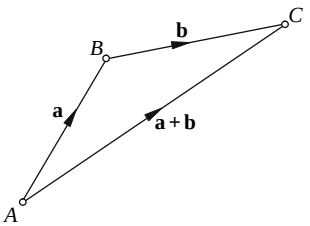


FIGURE 2.11.

From Fig. 2.11, we get

$$\mathbf{a} = \overrightarrow{AB} \quad \text{and} \quad \mathbf{b} = \overrightarrow{BC}$$

and

$$\mathbf{a} + \mathbf{b} = \overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC} \tag{2.1}$$

But equally, we could have described the vector addition shown in Fig. 2.12.

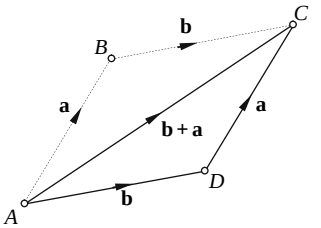


FIGURE 2.12.

From Fig. 2.12, we have

$$\mathbf{b} = \overrightarrow{AD} \quad \text{and } \mathbf{a} = \overrightarrow{DC}$$

and

$$\mathbf{b} + \mathbf{a} = \overrightarrow{AD} + \overrightarrow{DC} = \overrightarrow{AC}. \tag{2.2}$$

From Eqs. (2.1) and (2.2), we have

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}.$$

2.8.3 Problem 3

Let us now prove that the addition of vectors is associative, i.e., $\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}$.

Figure 2.13 shows four vectors forming a quadrilateral, where one diagonal is formed from $\mathbf{a} + \mathbf{b}$ and the other diagonal is formed from $\mathbf{b} + \mathbf{c}$.

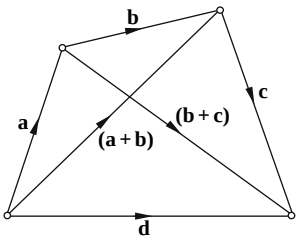


FIGURE 2.13.

It is obvious from Fig. 2.13 that

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = \mathbf{d}$$

and

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{d}$$

therefore,

$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c}.$$

2.9 Vector products

We now come to vector products, which originally took mathematicians of the day some time to recognise and define. What was unusual was that there were two ways of multiplying vectors together: one that produced a scalar quantity and another that created a new vector.

The product of two scalars is straightforward. For example, $3 \times 4 = 12$ and $4 \times 3 = 12$ — order has no impact on the final result. However, there are two interpretations of this product: one is simply the fact that 3 lots of 4 make 12; the other is that 3×4 can be thought of as an area, 12 units in size. The product of two vectors is not quite so simple, but there are similarities with scalar multiplication. For those readers who wish to know how the two products were discovered, see Michael Crowe's excellent book, *A History of Vector Analysis* [Crowe, 67].

2.9.1 Scalar product

Before we define this product, it's worth thinking about what we would have predicted the scalar product could have been based upon our knowledge of scalar multiplication. For example, given the following vectors:

$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}$$
$$\mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}$$

there is a temptation to multiply the corresponding vector components together:

$$x_a x_b, \quad y_a y_b, \quad z_a z_b.$$

But what do we do with these terms? Well, why not add them together? If we do, we get

$$\mathbf{a} \cdot \mathbf{b} = x_a x_b + y_a y_b + z_a z_b.$$

Well, it just so happens that this *is* the definition of the scalar product!

An alternative approach would be to multiply their lengths together: $\| \mathbf{a} \| \cdot \| \mathbf{b} \|$, which is a scalar. However, this ignores the orientation of the two vectors, which might tempt us to suggest one of the following extensions:

$$\| \mathbf{a} \| \cdot \| \mathbf{b} \| \cos\theta$$
$$\| \mathbf{a} \| \cdot \| \mathbf{b} \| \sin\theta$$
$$\| \mathbf{a} \| \cdot \| \mathbf{b} \| \theta$$

where θ is the separating angle between the vectors.

These all seem reasonable, and in mathematics there are no rules preventing anyone from defining some new formula or technique. What is important is that it integrates with the rest of mathematics. Well, as we are about to discover, the first two suggestions are extremely useful, but the third has no application.

The *scalar product* of two vectors $\mathbf{a} \cdot \mathbf{b}$ is defined as

$$\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \cdot \| \mathbf{b} \| \cos\theta \tag{2.3}$$

where

a and **b** are the two vectors

θ is the separating angle

and the dot '·' between **a** and **b** implies a scalar product.

The terms on the right-hand side (RHS) of the equals sign in Eq. (2.3) are scalars. Therefore, the result must also be a scalar — hence the name scalar product. The dot '·' between **a** and **b** has also given rise to a pseudonym for this operation: the *dot product*. Consequently, be prepared for both names being used for this operation.

Figure 2.14 shows three pairs of vectors with different separating angles. Notice that θ is the angle between the heads of the two vectors, and it never exceeds 180°.

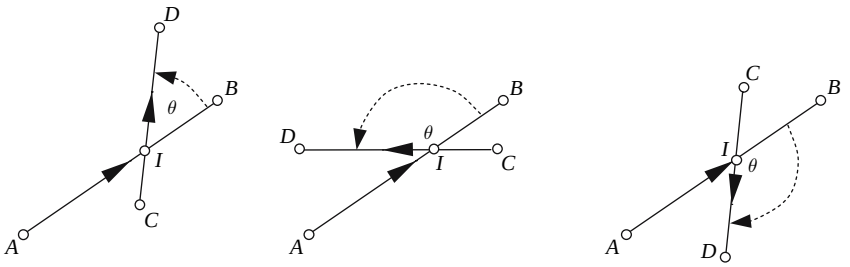


Figure 2.14.

The above definition is not so arbitrary as the preamble might suggest; it is extremely useful in a range of geometric and scientific applications. For example, if two vectors are perpendicular, their dot product is zero, because $\cos 90° = 0$. Furthermore, if two unit vectors have the same orientation, their dot product equals 1, because $\cos 0° = 1$. Such results are used as part of our problem-solving strategies later on.

Consider what happens when a force is applied to a mechanism restricted to move in one direction. Such a scenario is shown in Fig. 2.15.

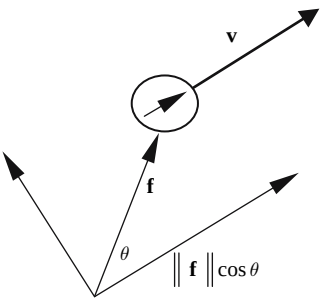


Figure 2.15.

If the force is represented by **f**, and the mechanism's direction by **v**, then the magnitude of the force acting in the direction of the mechanism is $\| \mathbf{f} \| \cos\theta$. And as the total work done is the product of the force over the acting distance, this is represented by $\| \mathbf{f} \| \| \mathbf{v} \| \cos\theta$, which is the dot product.

Figure 2.16 shows how the dot product should be visualized, where one vector is projected onto the other and the two lengths multiplied.

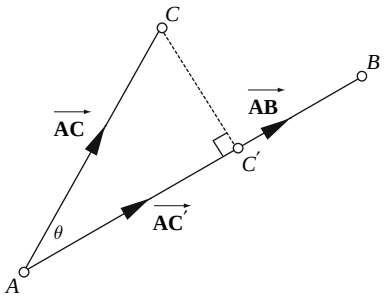


FIGURE 2.16.

The two vectors are $\overrightarrow{AB}$ and $\overrightarrow{AC}$ with a separating angle θ. The projection of $\overrightarrow{AC}$ onto $\overrightarrow{AB}$ is $\overrightarrow{AC'}$, which equals $\overrightarrow{AC}\cos\theta$. The dot product of $\overrightarrow{AB}$ and $\overrightarrow{AC}$ is therefore

$$\overrightarrow{AB} \cdot \overrightarrow{AC} = \left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{AC'} \right\| = \left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{AC} \right\| \cos\theta.$$

Let us show that

$$\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \| \mathbf{b} \| \cos\theta = x_a x_b + y_a y_b + z_a z_b.$$

We begin with

$$\mathbf{a} = x_a \mathbf{i} + y_a \mathbf{j} + z_a \mathbf{k}$$

and

$$\mathbf{b} = x_b \mathbf{i} + y_b \mathbf{j} + z_b \mathbf{k}.$$

Therefore,

$$\mathbf{a} \cdot \mathbf{b} = (x_a \mathbf{i} + y_a \mathbf{j} + z_a \mathbf{k}) \cdot (x_b \mathbf{i} + y_b \mathbf{j} + z_b \mathbf{k}).$$

Expanding

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} = {} & x_a x_b \mathbf{i} \cdot \mathbf{i} + y_a y_b \mathbf{j} \cdot \mathbf{j} + z_a z_b \mathbf{k} \cdot \mathbf{k} \\ & + x_a y_b \mathbf{i} \cdot \mathbf{j} + x_a z_b \mathbf{i} \cdot \mathbf{k} + y_a x_b \mathbf{j} \cdot \mathbf{i} + y_a z_b \mathbf{j} \cdot \mathbf{k} + z_a x_b \mathbf{k} \cdot \mathbf{i} + z_a y_b \mathbf{k} \cdot \mathbf{j}. \end{aligned}$$

We now need to discover the meaning of $\mathbf{i}\cdot\mathbf{i}, \mathbf{j}\cdot\mathbf{j}, \mathbf{k}\cdot\mathbf{k}, \mathbf{i}\cdot\mathbf{j}$,etc.

If we use Eq. (2.3) to evaluate $\mathbf{i}\cdot\mathbf{i}$, we get

$$\mathbf{i}\cdot\mathbf{i} = \|\,\mathbf{i}\,\|\,\|\,\mathbf{i}\,\|\cos 0^\circ = 1$$

The result is 1 because $\|\,\mathbf{i}\,\| = 1$, and the separating angle is 0°, whose cosine is 1. Obviously, this result also applies to $\mathbf{j}\cdot\mathbf{j}$ and $\mathbf{k}\cdot\mathbf{k}$. All the other vector combinations $\mathbf{i}\cdot\mathbf{j}, \mathbf{i}\cdot\mathbf{k}, \mathbf{j}\cdot\mathbf{k}$, etc., have a separating angle of 90°, whose cosine is zero. Consequently, all of these terms vanish and we are left with

$$\mathbf{a}\cdot\mathbf{b} = \|\,\mathbf{a}\,\|\cdot\|\,\mathbf{b}\,\|\cos\theta = x_a x_b + y_a y_b + z_a z_b \tag{2.4}$$

which is the full definition of the scalar or dot product and should be committed to memory.

Normally, we do not know the value of θ, and Eq. (2.4) is used to discover its value using the following equation:

$$\theta = \cos^{-1}\left(\frac{x_a x_b + y_a y_b + z_a z_b}{\|\,\mathbf{a}\,\|\,\|\,\mathbf{b}\,\|}\right) \tag{2.5}$$

Note that if $\mathbf{a}$ and $\mathbf{b}$ are unit vectors, then

$$\theta = \cos^{-1}\left(x_a x_b + y_a y_b + z_a z_b\right)$$

Note also that Eq. (2.5) provides an unsigned angle, θ. But we know that when working in 2D, clockwise angles are negative, while counter-clockwise angles are positive. Well, the sign of θ is available for 2D angles and is determined by the *perp product* in Section 2.10.

Before proceeding, let's test Eqs. (2.4) and (2.5) with two examples:

Example 1

Find the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$, given

$$\mathbf{a} = \mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$$

and

$$\mathbf{b} = 4\mathbf{i} + 5\mathbf{j} + 6\mathbf{k}.$$

Therefore,

$$\|\,\mathbf{a}\,\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

and

$$\|\,\mathbf{b}\,\| = \sqrt{4^2 + 5^2 + 6^2} = \sqrt{77}.$$

Using Eq. (2.4), we find

$$\mathbf{a}\cdot\mathbf{b} = \sqrt{14}\sqrt{77}\cos\theta = 1\times 4 + 2\times 5 + 3\times 6 = 32.$$

Using Eq. (2.5), we get

$$\theta = \cos^{-1}\left(\frac{32}{\sqrt{14}\sqrt{77}}\right) = 12.9^\circ.$$

Example 2

Prove that two vectors are perpendicular, given that

$$\mathbf{a} = \mathbf{i} + 3\mathbf{j} - 2\mathbf{k}$$

and

$$\mathbf{b} = 4\mathbf{i} + 2\mathbf{j} + 5\mathbf{k}$$

Therefore,

$$\mathbf{a} \cdot \mathbf{b} = (\mathbf{i} + 3\mathbf{j} - 2\mathbf{k}) \cdot (4\mathbf{i} + 2\mathbf{j} + 5\mathbf{k}) = 0.$$

Because $\| \mathbf{a} \| > 0$ and $\| \mathbf{b} \| > 0$ and the dot product is zero, it can only mean that the separating angle of the vectors is 90°, whose cosine is zero.

Having looked at two examples, let's return to the geometry behind the dot product.

Figure 2.16 shows that

$$\overrightarrow{AB} \cdot \overrightarrow{AC} = \left\| \overrightarrow{AB} \right\| \cdot \left\| \overrightarrow{AC} \right\| \cos\theta = \left\| \overrightarrow{AB} \right\| \cdot \left\| \overrightarrow{AC'} \right\|.$$

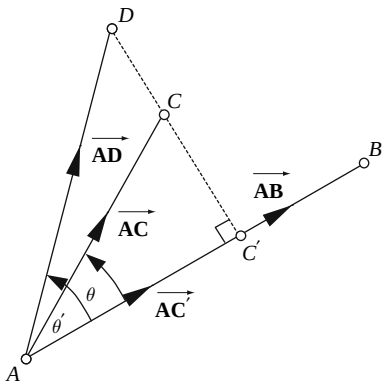


Figure 2.17.

But Fig. 2.17 introduces a point D such that the line $\overrightarrow{C'CD}$ is perpendicular to $\overrightarrow{AB}$. Surely, the projection of $\overrightarrow{AD}$ on $\overrightarrow{AB}$ is $\overrightarrow{AC'}$, which equals the projection of $\overrightarrow{AC}$ on $\overrightarrow{AB}$? In fact, all points on a line perpendicular to another line will have a common projection.

Therefore,

$$\overrightarrow{AB} \cdot \overrightarrow{AD} = \left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{AD} \right\| \cos\theta' = \left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{AC'} \right\|.$$

This geometric configuration often arises when solving problems. For example, with reference to Fig. 2.18, we can state directly that

$$\mathbf{p} \cdot \mathbf{v} = \mathbf{q} \cdot \mathbf{v}$$

simply because $\mathbf{p}$ and $\mathbf{q}$ have identical projections on $\mathbf{v}$.

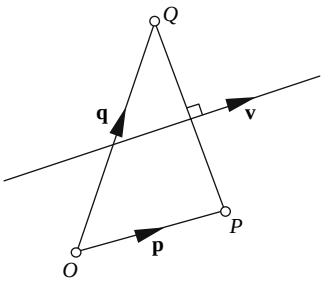


FIGURE 2.18.

Now let's take a look at the second way of multiplying vectors – the vector product.

2.9.2 Vector product

Again, before defining this product, let's try to anticipate what the answer could be. We start with two 3D vectors, **a** and **b**, which must lie in some common plane in space. Just in case you find this hard to imagine, remember that vectors can be moved around in space as long as their length and orientation with the axial system are not disturbed. So, if we bring two such vectors together, as shown in Fig. 2.19, they will share a common plane.

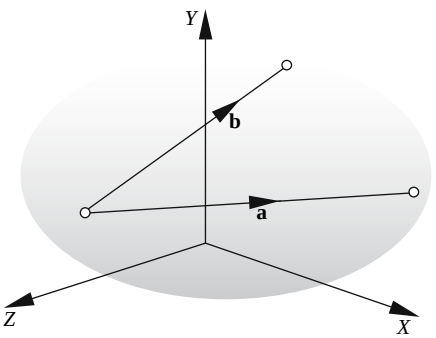


FIGURE 2.19.

Now, if we multiply these vectors together and create a third vector, it is obvious that the vector must either be parallel to or intersect the common plane. But wouldn't it be nice if the vector were perpendicular to the plane? Well, it just so happens that this is the case!

Back in the 19th century, Sir William Rowan Hamilton was struggling to discover a 3D equivalent of complex numbers. Given that a 2D complex number is expressed as

$$q = s + ai$$

it seemed reasonable to conjecture that a 3D complex number would be of the form

$$q = s + ai + bj$$

where i and j equal $\sqrt{-1}$. But when two such objects are multiplied together, they create terms such as $i \cdot i$, $j \cdot j$, $i \cdot j$, and $j \cdot i$. The first two are no problem as they are equal to -1, but the last two did present a problem, and there was no obvious solution, so he extended the notation to

$$q = s + ai + bj + ck$$

where i, j, and k equal $\sqrt{-1}$. When multiplying these objects together, Hamilton ended up with terms such as $i \cdot j, j \cdot k$, and $k \cdot i$, which also defied description, until October 16, 1843, when he thought of the idea that $i \cdot j = k$, $j \cdot k = i$, and $k \cdot i = j$. He also conjectured that $j \cdot i = -k$, $k \cdot j = -i$, and $i \cdot k = -j$. On that day, he discovered *quaternions*, which revealed the foundations of the scalar and vector products.

Hamilton's i, j, and k are not the $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ we have been using, but there are some strange similarities. His vector product suggested the following manipulation:

$$\mathbf{a} \times \mathbf{b} = (y_a z_b - y_b z_a)\,\mathbf{i} + (z_a x_b - z_b x_a)\,\mathbf{j} + (x_a y_b - x_b y_a)\,\mathbf{k} \tag{2.6}$$

which, clearly, is another vector and turns out to be perpendicular to the plane containing both $\mathbf{a}$ and $\mathbf{b}$. Equation (2.6) can also be expressed in its determinant form:

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} y_a & z_a \\ y_b & z_b \end{vmatrix} \mathbf{i} + \begin{vmatrix} z_a & x_a \\ z_b & x_b \end{vmatrix} \mathbf{j} + \begin{vmatrix} x_a & y_a \\ x_b & y_b \end{vmatrix} \mathbf{k} \tag{2.7}$$

and is often shown as

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_a & y_a & z_a \\ x_b & y_b & z_b \end{vmatrix} \tag{2.8}$$

which produces the same result as Eq. (2.7) and is easier to remember. Note that the '$\times$' symbol is used to distinguish this product from the scalar product, and like the scalar product has a pseudonym: the *cross product*.

Let's approach this product from another way. Given

$$\mathbf{a} = x_a \mathbf{i} + y_a \mathbf{j} + z_a \mathbf{k}$$

and

$$\mathbf{b} = x_b \mathbf{i} + y_b \mathbf{j} + z_b \mathbf{k}$$

then

$$\mathbf{a} \times \mathbf{b} = (x_a \mathbf{i} + y_a \mathbf{j} + z_a \mathbf{k}) \times (x_b \mathbf{i} + y_b \mathbf{j} + z_b \mathbf{k})$$

which expands to

$$\begin{aligned} \mathbf{a} \times \mathbf{b} = {} & x_a x_b \mathbf{i} \times \mathbf{i} + y_a y_b \mathbf{j} \times \mathbf{j} + z_a z_b \mathbf{k} \times \mathbf{k} \\ & + x_a y_b \mathbf{i} \times \mathbf{j} + x_a z_b \mathbf{i} \times \mathbf{k} + y_a x_b \mathbf{j} \times \mathbf{i} + y_a z_b \mathbf{j} \times \mathbf{k} + z_a x_b \mathbf{k} \times \mathbf{i} + z_a y_b \mathbf{k} \times \mathbf{j}. \end{aligned} \tag{2.9}$$

If you look carefully at Eq. (2.9), you will see features of Eq. (2.6). Now we want a rule similar to Eq. (2.3) as that used for the scalar product. So let's try the following:

$$\mathbf{a} \times \mathbf{b} = \| \mathbf{a} \| \| \mathbf{b} \| \sin\theta \, \hat{\mathbf{n}} \tag{2.10}$$

where the unit vector $\hat{\mathbf{n}}$ is perpendicular to both **a** and **b**.

Applying Eq. (2.10) to Eq. (2.9), the terms $\mathbf{i} \times \mathbf{i}$, $\mathbf{j} \times \mathbf{j}$, and $\mathbf{k} \times \mathbf{k}$ collapse to zero, because the separating angle is 0°, and $\sin 0° = 0$. So we are left with

$$\mathbf{a} \times \mathbf{b} = x_a y_b \mathbf{i} \times \mathbf{j} + x_a z_b \mathbf{i} \times \mathbf{k} + y_a x_b \mathbf{j} \times \mathbf{i} + y_a z_b \mathbf{j} \times \mathbf{k} + z_a x_b \mathbf{k} \times \mathbf{i} + z_a y_b \mathbf{k} \times \mathbf{j}$$

If we invoke Hamilton's rules for quaternions for vectors:

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \mathbf{j} \times \mathbf{k} = \mathbf{i} \text{ and } \mathbf{k} \times \mathbf{i} = \mathbf{j}$$

and their reverse forms:

$$\mathbf{j} \times \mathbf{i} = -\mathbf{k}, \mathbf{k} \times \mathbf{j} = -\mathbf{i} \text{ and } \mathbf{i} \times \mathbf{k} = -\mathbf{j}$$

we obtain

$$\mathbf{a} \times \mathbf{b} = x_a y_b \mathbf{k} - x_a z_b \mathbf{j} - y_a x_b \mathbf{k} + y_a z_b \mathbf{i} + z_a x_b \mathbf{j} - z_a y_b \mathbf{i}.$$

Reordering the terms gives

$$\mathbf{a} \times \mathbf{b} = y_a z_b \mathbf{i} - z_a y_b \mathbf{i} + z_a x_b \mathbf{j} - x_a z_b \mathbf{j} + x_a y_b \mathbf{k} - y_a x_b \mathbf{k}$$

and, finally,

$$\mathbf{a} \times \mathbf{b} = (y_a z_b - z_a y_b)\,\mathbf{i} + (z_a x_b - x_a z_b)\,\mathbf{j} + (x_a y_b - y_a x_b)\,\mathbf{k}$$

which is identical to Eq. (2.6)!

All of this may appear longwinded, but it is necessary to bring to life what happened historically. A mathematician suddenly thought of a formula for the vector product.

Today, math books contain statements such as, "The **vector product** is defined as

$$\mathbf{a} \times \mathbf{b} = \mathbf{n} = \| \mathbf{a} \| \| \mathbf{b} \| \sin\theta \, \hat{\mathbf{n}} \tag{2.11}$$

where

θ is the smaller angle between **a** and **b**

and

n is perpendicular to both **a** and **b**".

Also,

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_a & y_a & z_a \\ x_b & y_b & z_b \end{vmatrix} \tag{2.12}$$

These definitions are true, but they hide the agony that went on behind the scenes at the time.

Taking Eq. (2.11) as our definition, let's consider the orientation of **n**.

As we are employing a right-handed system of axes, our right hand also determines the orientation of **n**. Therefore, with reference to our right hand, the thumb represents **a**, the first finger represents **b**, and the middle finger provides the direction of **n**. It is convenient to think of this operation as a rotation from **a** to **b** producing the perpendicular vector **n**. This is shown in Fig. 2.20. Note that the vector product is very sensitive to the order of the two vectors. Switch the vectors around and we get

$$\mathbf{b} \times \mathbf{a} = -\mathbf{n} = -\| \mathbf{b} \| \| \mathbf{a} \| \sin\theta \hat{\mathbf{n}}$$

and we see that the vector product breaks the commutative rule we take for granted with scalars. In fact, $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$.

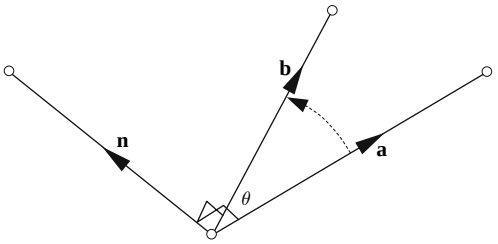


Figure 2.20.

Equation (2.11) not only determines the length of **n**, but also says something about the space enclosed by **a** and **b**. And if we remove $\hat{\mathbf{n}}$ from the right-hand side, we obtain $\| \mathbf{a} \| \| \mathbf{b} \| \sin\theta$, which may remind you of another formula: $\frac{1}{2}ab\sin\theta$, which is the area of a triangle with sides a and b with an enclosed angle θ. Figure 2.21 illustrates this.

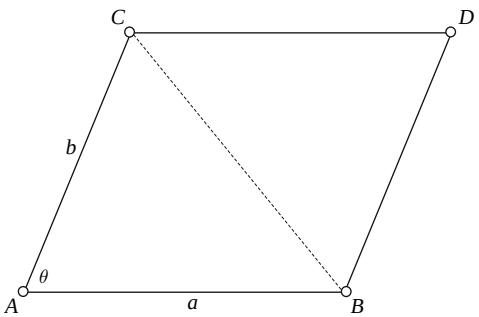


Figure 2.21.

The area of ΔABC equals $\frac{1}{2}ab\sin\theta$, which means that $ab\sin\theta$ equals the area of the parallelogram $ABDC$. Thus, when we compute the cross product $\mathbf{a} \times \mathbf{b}$, the perpendicular vector **n** has a length equal to the area of the parallelogram with sides $\| \mathbf{a} \|$ and $\| \mathbf{b} \|$. Figure 2.22 illustrates this relationship.

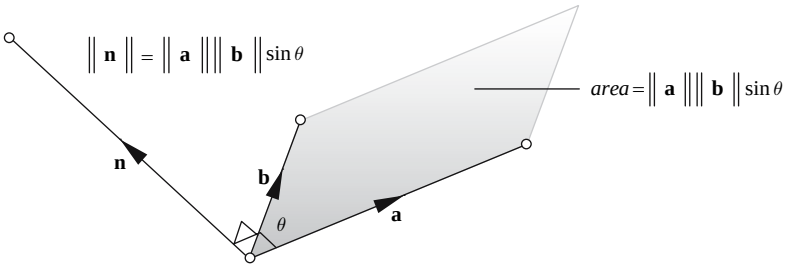


FIGURE 2.22.

Before proceeding, it will be useful to prove that the above formulas are consistent, so let's demonstrate that the rectangular unit vectors obey these rules.

Substituting **i** and **j** in Eq. (2.12), we get

$$\mathbf{i}\times\mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix}$$

$$\mathbf{i}\times\mathbf{j} = \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}\mathbf{i} + \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}\mathbf{k} = \mathbf{k}.$$

Substituting **j** and **k** in Eq. (2.12), we get

$$\mathbf{j}\times\mathbf{k} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

$$\mathbf{j}\times\mathbf{k} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}\mathbf{i} + \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}\mathbf{k} = \mathbf{i}.$$

Substituting **k** and **i** in Eq. (2.12), we get

$$\mathbf{k}\times\mathbf{i} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix}$$

$$\mathbf{k}\times\mathbf{i} = \begin{vmatrix} 0 & 1 \\ 0 & 0 \end{vmatrix}\mathbf{i} + \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 0 & 0 \\ 1 & 0 \end{vmatrix}\mathbf{k} = \mathbf{j}$$

both of which are correct. Now let's show how the vector product is sensitive to vector sequence by reversing one of the above: substituting **k** and **j** in Eq. (2.12), we get

$$\mathbf{k}\times\mathbf{j} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

$$\mathbf{k}\times\mathbf{j}=\begin{vmatrix}0&1\\1&0\end{vmatrix}\mathbf{i}+\begin{vmatrix}1&0\\0&0\end{vmatrix}\mathbf{j}+\begin{vmatrix}0&0\\0&1\end{vmatrix}\mathbf{k}=-\mathbf{i}$$

which confirms that

$$\mathbf{j}\times\mathbf{k}=-\mathbf{k}\times\mathbf{j}.$$

The areas of the parallelograms associated with the above examples all equal unity, which is the length of the resulting vector.

Perhaps now you can understand how Sir William Rowan Hamilton must have felt when he discovered the above!

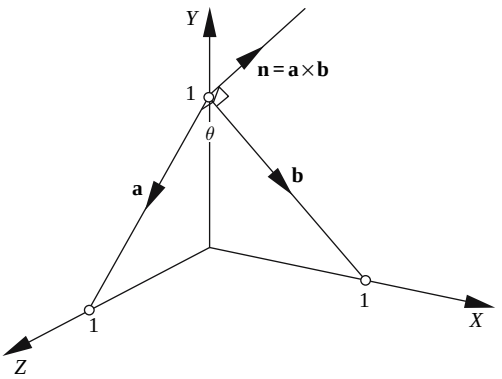


FIGURE 2.23.

As a further example, consider the scenario shown in Fig. 2.23. Vectors **a** and **b** are given by

$$\mathbf{a}=-\mathbf{j}+\mathbf{k}$$

and

$$\mathbf{b}=\mathbf{i}-\mathbf{j}.$$

The task is to compute the vectors normal to both **a** and **b**. Therefore,

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\0&-1&1\\1&-1&0\end{vmatrix}$$

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}-1&1\\-1&0\end{vmatrix}\mathbf{i}+\begin{vmatrix}1&0\\0&1\end{vmatrix}\mathbf{j}+\begin{vmatrix}0&-1\\1&-1\end{vmatrix}\mathbf{k}$$

$$\mathbf{a}\times\mathbf{b}=\mathbf{i}+\mathbf{j}+\mathbf{k}$$

which are correct. But what about the angle θ? Well, this can be computed by Eq. (2.5) or by rearranging Eq. (2.11). Using Eq. (2.5), we get

$$\theta=\cos^{-1}\left(\frac{x_a x_b+y_a y_b+z_a z_b}{\|\mathbf{a}\|\,\|\mathbf{b}\|}\right)$$

$$\theta = \cos^{-1}\left(\frac{0\times 1+(-1)\times(-1)+1\times 0}{\sqrt{2}\sqrt{2}}\right) = 60°.$$

Using Eq. (2.11), we get

$$\theta = \sin^{-1}\left(\frac{\|\,\mathbf{n}\,\|}{\|\,\mathbf{a}\,\|\,\|\,\mathbf{b}\,\|}\right) = \sin^{-1}\left(\frac{\sqrt{3}}{\sqrt{2}\sqrt{2}}\right) = 60°.$$

The preceding three equations are correct, as the triangle formed by **a** and **b** is an equilateral triangle.

2.9.3 Surface normals

One important application of the vector product in computer graphics is its ability to compute *surface normals*. Given a triangular mesh, we can use two sides of any triangle as the reference vectors for the vector product. However, we must know the convention used for defining the vertices of the triangle, and which side of the triangle faces outside or inside. Figure 2.24 shows such a triangle where its vertices are defined in a counter-clockwise sequence as viewed from the outside.

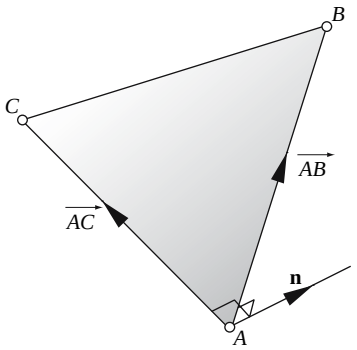


Figure 2.24.

Using our right hand, we can see that first vector has to be $\overrightarrow{AB}$ followed by $\overrightarrow{AC}$, which ensures that the surface normal **n** points in the direction from where we are viewing the triangle.

Given that the vertices are $A\,(x_A, y_A, z_A)$, $B\,(x_B, y_B, z_B)$, and $C\,(x_C, y_C, z_C)$, then the vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are defined as follows:

$$\overrightarrow{AB} = (x_B - x_A)\,\mathbf{i} + (y_B - y_A)\,\mathbf{j} + (z_B - z_A)\,\mathbf{k}$$

$$\overrightarrow{AC} = (x_C - x_A)\,\mathbf{i} + (y_C - y_A)\,\mathbf{j} + (z_C - z_A)\,\mathbf{k}.$$

Therefore,

$$\mathbf{n} = \overrightarrow{AB}\times\overrightarrow{AC} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_B - x_A & y_B - y_A & z_B - z_A \\ x_C - x_A & y_C - y_A & z_C - z_A \end{vmatrix}.$$

2.9.4 The algebra of vector products

Before proceeding with triple products, we need to consolidate the algebraic rules associated with the dot and cross products. Two laws need to be considered: the commutative and distributive laws, but we also need to confirm the role of scalars in these products.

Commutative law $\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a}$

As

$$\mathbf{a}\cdot\mathbf{b}=\|\mathbf{a}\|\,\|\mathbf{b}\|\cos\theta \tag{2.13}$$

the commutative law of scalar multiplication permits us to write Eq. (2.13) as

$$\mathbf{a}\cdot\mathbf{b}=\|\mathbf{b}\|\,\|\mathbf{a}\|\cos\theta=\mathbf{b}\cdot\mathbf{a}$$

Therefore,

$$\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a}.$$

Distributive law $\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}$

Given

$$\mathbf{a}=x_a\mathbf{i}+y_a\mathbf{j}+z_a\mathbf{k}$$
$$\mathbf{b}=x_b\mathbf{i}+y_b\mathbf{j}+z_b\mathbf{k}$$
$$\mathbf{c}=x_c\mathbf{i}+y_c\mathbf{j}+z_c\mathbf{k}$$

then

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=(x_a\mathbf{i}+y_a\mathbf{j}+z_a\mathbf{k})\cdot[(x_b\mathbf{i}+y_b\mathbf{j}+z_b\mathbf{k})+(x_c\mathbf{i}+y_c\mathbf{j}+z_c\mathbf{k})]$$
$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=(x_a\mathbf{i}+y_a\mathbf{j}+z_a\mathbf{k})\cdot[(x_b+x_c)\,\mathbf{i}+(y_b+y_c)\,\mathbf{j}+(z_b+z_c)\,\mathbf{k}]$$
$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=x_a\,(x_b+x_c)+y_a\,(y_b+y_c)+z_a\,(z_b+z_c)$$
$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=(x_ax_b+x_ax_c+y_ay_b)+(y_ay_c+z_az_b+z_az_c)=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}$$

Therefore,

$$\mathbf{a}\cdot(\mathbf{b}+\mathbf{c})=\mathbf{a}\cdot\mathbf{b}+\mathbf{a}\cdot\mathbf{c}.$$

Commutative law $\mathbf{a}\times\mathbf{b}\neq\mathbf{b}\times\mathbf{a}$

The commutative law does not hold for the vector product.

Given

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\x_a&y_a&z_a\\x_b&y_b&z_b\end{vmatrix}$$

$$\mathbf{a}\times\mathbf{b}=\begin{vmatrix}y_a&z_a\\y_b&z_b\end{vmatrix}\mathbf{i}+\begin{vmatrix}z_a&x_a\\z_b&x_b\end{vmatrix}\mathbf{j}+\begin{vmatrix}x_a&y_a\\x_b&y_b\end{vmatrix}\mathbf{k}$$

$$\mathbf{a}\times\mathbf{b}=(y_az_b-y_bz_a)\,\mathbf{i}+(z_ax_b-z_bx_a)\,\mathbf{j}+(x_ay_b-x_by_a)\,\mathbf{k}.$$

In contrast,

$$\mathbf{b}\times\mathbf{a}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\x_b&y_b&z_b\\x_a&y_a&z_a\end{vmatrix}$$

$$\mathbf{b}\times\mathbf{a}=\begin{vmatrix}y_b&z_b\\y_a&z_a\end{vmatrix}\mathbf{i}+\begin{vmatrix}z_b&x_b\\z_a&x_a\end{vmatrix}\mathbf{j}+\begin{vmatrix}x_b&y_b\\x_a&y_a\end{vmatrix}\mathbf{k}$$

$$\mathbf{b}\times\mathbf{a}=(y_bz_a-y_az_b)\,\mathbf{i}+(z_bx_a-z_ax_b)\,\mathbf{j}+(x_by_a-x_ay_b)\,\mathbf{k}$$

$$\mathbf{b}\times\mathbf{a}=-(y_az_b-y_bz_a)\,\mathbf{i}+(z_ax_b-z_bx_a)\,\mathbf{j}+(x_ay_b-x_by_a)\,\mathbf{k}=-\mathbf{a}\times\mathbf{b}$$

Therefore,

$$\mathbf{a}\times\mathbf{b}=-\mathbf{b}\times\mathbf{a}.$$

Distributive law $\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}$

Given

$$\mathbf{a}=x_a\mathbf{i}+y_a\mathbf{j}+z_a\mathbf{k}$$

$$\mathbf{b}=x_b\mathbf{i}+y_b\mathbf{j}+z_b\mathbf{k}$$

$$\mathbf{c}=x_c\mathbf{i}+y_c\mathbf{j}+z_c\mathbf{k}$$

and

$$\mathbf{d}=\mathbf{b}+\mathbf{c}=x_d\mathbf{i}+y_d\mathbf{j}+z_d\mathbf{k}$$

Then

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c})=\mathbf{a}\times\mathbf{d}=\begin{vmatrix}\mathbf{i}&\mathbf{j}&\mathbf{k}\\x_a&y_a&z_a\\x_d&y_d&z_d\end{vmatrix}$$

$$\mathbf{a}\times\mathbf{d}=\begin{vmatrix}y_a&z_a\\y_d&z_d\end{vmatrix}\mathbf{i}+\begin{vmatrix}z_a&x_a\\z_d&x_d\end{vmatrix}\mathbf{j}+\begin{vmatrix}x_a&y_a\\x_d&y_d\end{vmatrix}\mathbf{k}.$$

Substituting $\mathbf{b}+\mathbf{c}$ for $\mathbf{d}$, we get

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \begin{vmatrix} y_a & z_a \\ y_b+y_c & z_b+z_c \end{vmatrix}\mathbf{i} + \begin{vmatrix} z_a & x_a \\ z_b+z_c & x_b+x_c \end{vmatrix}\mathbf{j} + \begin{vmatrix} x_a & y_a \\ x_b+x_c & y_b+y_c \end{vmatrix}\mathbf{k}$$

$$\begin{aligned}
\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= [y_a(z_b+z_c)-z_a(y_b+y_c)]\,\mathbf{i} \\
&\quad +[z_a(x_b+x_c)-x_a(z_b+z_c)]\,\mathbf{j} \\
&\quad +[x_a(y_b+y_c)-y_a(x_b+x_c)]\,\mathbf{k} \\
\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= (y_a z_b - y_b z_a)\,\mathbf{i} + (y_a z_c - y_c z_a)\,\mathbf{i} \\
&\quad +(z_a x_b - z_b x_a)\,\mathbf{j} + (z_a x_c - z_c x_a)\,\mathbf{j} \\
&\quad +(x_a y_b - x_b y_a)\,\mathbf{k} + (x_a y_c - x_c y_a)\,\mathbf{k} \\
\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}.
\end{aligned}$$

2.10 Triple products

Having examined the products that involve two vectors, now let's look at the products that involve three vectors. An interesting approach is to take three vectors, **a**, **b**, **c**, and explore the combinations using a '·' and a '×' without disturbing the alphabetic order of the vector names. These are summarised in Table 2.3.

Table 2.3

$\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c})$	$\mathbf{a}\times(\mathbf{b}\cdot\mathbf{c})$
$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$	$(\mathbf{a}\cdot\mathbf{b})\times\mathbf{c}$
$\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$	$\mathbf{a}\cdot(\mathbf{b}\cdot\mathbf{c})$
$(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$	$(\mathbf{a}\cdot\mathbf{b})\cdot\mathbf{c}$

Not all the combinations in Table 2.3 lead to a meaningful result, and for neatness the meaningless combinations are arranged in the second column. They are meaningless because the operation in parentheses is a dot product, which is a scalar, and cannot be associated with a scalar product or vector product. So we are left with the four triple products in the first column, which do satisfy the rules associated with the scalar and vector products. But what do they mean? Well, let's find out.

2.10.1 Scalar triple product

Figure 2.25 shows three vectors, **a**, **b**, and **c**, that provide the basis for a prism, where the area of the base is given by $Area = \|\mathbf{a}\|\,\|\mathbf{b}\|\sin\theta$. But $\mathbf{n}=\mathbf{a}\times\mathbf{b}$, where $\|\mathbf{n}\| = Area$. Now the volume of a prism is the product of its (vertical height)·(base area), which in vector notation is given by

$$Volume = (\|\mathbf{c}\|\cos\alpha)\cdot(\|\mathbf{n}\|) = \|\mathbf{c}\|\,\|\mathbf{n}\|\cos\alpha$$

which we recognise as the dot product $\mathbf{c} \cdot \mathbf{n}$. But as $\mathbf{n} = \mathbf{a} \times \mathbf{b}$, we can state

$$Volume = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$$

which is a rather elegant relationship.

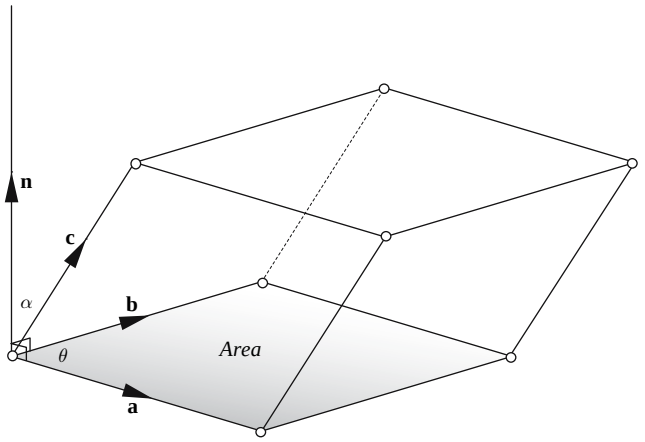


FIGURE 2.25.

The prism in Fig. 2.25 could easily be drawn with vectors **b** and **c** as the base, which, using the above reasoning, would make

$$Volume = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) .$$

Similarly, if vectors **c** and **a** form the base, the volume is

$$Volume = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a})$$

thus, we can conclude that

$$Volume = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) . \tag{2.14}$$

The first two entries in the first column of Table 2.3 correspond to the volume calculating qualities of this *scalar triple product*.

Unlike the vector product, the dot product is commutative, which permits us to write

$$\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$$

which, using Eq. (2.14), leads to

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} \tag{2.15}$$

and as Eq. (2.15) only makes sense when the parentheses embrace the cross product, we employ $[\mathbf{abc}]$ to represent $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ or $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$, which is also called the *box product*. However, the

cross product is not commutative, which means that if the vectors representing the base become reversed, the resulting normal vector **n** points in the opposite direction to the third vector, which in turn makes the final dot product negative. Thus, strictly speaking, [**abc**] represents the signed volume of the prism. To illustrate this, we will compute the volume of a cube and show why vector order is so important.

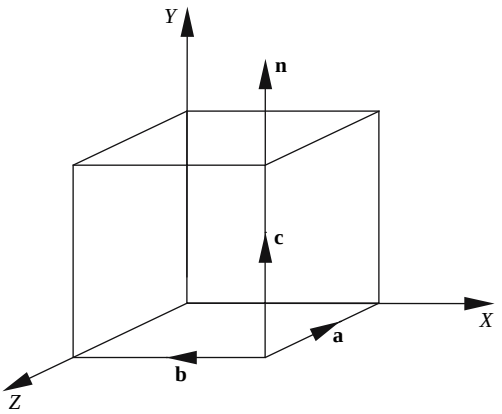


FIGURE 2.26.

Figure 2.26 shows three vectors, **a**, **b**, and **c**, that provide the basis for a cube. The vectors are defined as

$$\mathbf{a} = -\mathbf{k}$$

$$\mathbf{b} = -\mathbf{i}$$

$$\mathbf{c} = \mathbf{j}$$

and are organised in a right-hand sequence.

Substituting these in

$$[\mathbf{abc}] = Volume = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} \tag{2.16}$$

we obtain

$$[\mathbf{abc}] = (-\mathbf{k} \times -\mathbf{i}) \cdot \mathbf{j}.$$

First, let's compute the cross product

$$(-\mathbf{k} \times -\mathbf{i})$$

$$(-\mathbf{k} \times -\mathbf{i}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{vmatrix} = \mathbf{j}.$$

Now compute the dot product

$$\mathbf{j} \cdot \mathbf{j} = +1.$$

Thus, the cube has a volume of +1 units.

Now let's interchange the base vectors so that

$$\mathbf{a} = -\mathbf{i}$$

$$\mathbf{b} = -\mathbf{k}$$

$$\mathbf{c} = \mathbf{j}$$

which are organised in a left-hand sequence.

Substituting these in Eq. (2.16), we obtain

$$[\mathbf{abc}] = (-\mathbf{i} \times -\mathbf{k}) \cdot \mathbf{j}$$

Let's compute the cross product:

$$(-\mathbf{i} \times -\mathbf{k})$$

$$(-\mathbf{i} \times -\mathbf{k}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{vmatrix} = -\mathbf{j}$$

Now compute the dot product :

$$-\mathbf{j} \cdot \mathbf{j} = -1 \text{ units.}$$

Thus, the cube has a volume of −1 units, and we see that

$$[\mathbf{abc}] = \textit{Volume} = |\,(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}\,| = |\,\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})\,|\,. \tag{2.17}$$

We can compact the above calculation as follows: given

$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}$$

$$\mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}$$

$$\mathbf{c} = x_c\mathbf{i} + y_c\mathbf{j} + z_c\mathbf{k}$$

then

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{a} \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix}$$

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}) \cdot \begin{bmatrix} (y_b z_c - y_c z_b)\,\mathbf{i} + \\ (z_b x_c - z_c x_b)\,\mathbf{j} + \\ (x_b y_c - x_c y_b)\,\mathbf{k} \end{bmatrix}$$

and

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = x_a\,(y_b z_c - y_c z_b) + y_a\,(z_b x_c - z_c x_b) + z_a\,(x_b y_c - x_c y_b)$$

which equals

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} \tag{2.18}$$

which is also an elegant relationship.

Furthermore, if two columns or rows of a determinant are interchanged, its sign is reversed. Also, the value of a determinant is unaffected if its elements are transposed. Therefore,

$$\begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} = - \begin{vmatrix} x_a & y_a & z_a \\ x_c & y_c & z_c \\ x_b & y_b & z_b \end{vmatrix} = - \begin{vmatrix} x_b & y_b & z_b \\ x_a & y_a & z_a \\ x_c & y_c & z_c \end{vmatrix}$$

and

$$[\mathbf{abc}] = - \begin{vmatrix} \mathbf{a} & \mathbf{c} & \mathbf{b} \end{vmatrix} = -\mathbf{a} \cdot (\mathbf{c} \times \mathbf{b}) \cdot = -\mathbf{b} \cdot (\mathbf{a} \times \mathbf{c}) .$$

Just to prove that Eq. (2.18) works, let's take the first example above, where

$$\mathbf{a} = -\mathbf{k}$$

$$\mathbf{b} = -\mathbf{i}$$

$$\mathbf{c} = \mathbf{j}$$

Using Eq. (2.18), we find

$$[\mathbf{abc}] = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = 1$$

We are now in a position to prove that $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$, even though we concluded it in Eq. (2.14).

From Eq. (2.18), we have

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} .$$

It can be shown that exchanging two rows of a determinant changes its sign. Therefore,

$$\begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} = - \begin{vmatrix} x_b & y_b & z_b \\ x_a & y_a & z_a \\ x_c & y_c & z_c \end{vmatrix} = \begin{vmatrix} x_b & y_b & z_b \\ x_c & y_c & z_c \\ x_a & y_a & z_a \end{vmatrix} = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) .$$

Similarly,

$$\begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} = - \begin{vmatrix} x_c & y_c & z_c \\ x_b & y_b & z_b \\ x_a & y_a & z_a \end{vmatrix} = \begin{vmatrix} x_c & y_c & z_c \\ x_a & y_a & z_a \\ x_b & y_b & z_b \end{vmatrix} = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) .$$

Therefore,

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) . \tag{2.19}$$

2.10.2 The vector triple product

One of the vector triple products in Table 2.3 is $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$, and there are three questions we can ask about it:

1. Is $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ equivalent to an expression involving dot products?
2. Does $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ have any connection with $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$?
3. Does $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ have any geometric meaning?

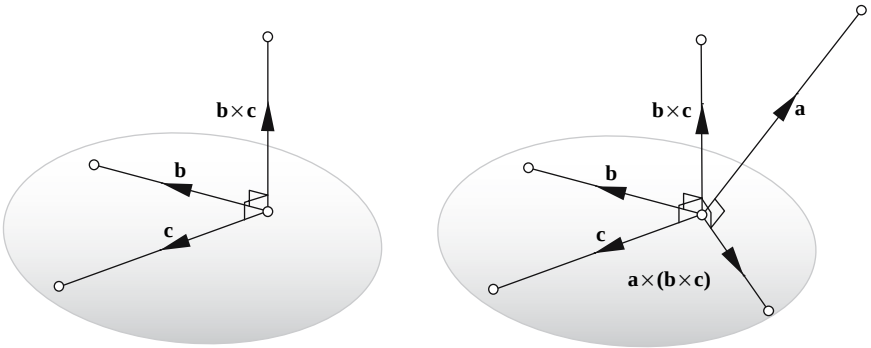


FIGURE 2.27.

Figure 2.27(a) shows part of the triple product, where we see vectors $\mathbf{b}$ and $\mathbf{c}$ sharing a common plane and the cross product $\mathbf{b} \times \mathbf{c}$ perpendicular to this plane. The third vector $\mathbf{a}$ can either reside in this plane or intersect it. First, let's consider the scenario where it intersects the plane, as shown in Fig. 2.27(b).

If we take the cross product of $\mathbf{a}$ with $\mathbf{b} \times \mathbf{c}$, the result $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ must be perpendicular to $\mathbf{a}$ and $\mathbf{b} \times \mathbf{c}$, which means that it must be in the original plane containing $\mathbf{b}$ and $\mathbf{c}$. If you find this difficult to visualise, don't worry. Try this: if a vector exists perpendicular to some plane, and we have to compute a vector product with it and some other vector, the result of this product must be perpendicular to the original vector, which must be in the perpendicular plane! Thus, we see in Fig. 2.27(b) the result of $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ lying in the plane containing $\mathbf{b}$ and $\mathbf{c}$.

Now let's consider the scenario where $\mathbf{a}$ resides in the plane containing $\mathbf{b}$ and $\mathbf{c}$. Well, if we think about it, it is no different from the previous scenario, because the result $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ must be perpendicular to $\mathbf{b} \times \mathbf{c}$, which is the plane! The same is true for $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$.

It can be shown that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c} \tag{2.20}$$

and

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\,\mathbf{a}. \tag{2.21}$$

It is obvious that $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$, which confirms that the associative law does not hold for the vector triple product. So the answer to the second question we raised at the beginning of this section is no.

From what we have discovered above, there does not seem to be any geometric significance to the vector triple product, so the answer to the third question is also no.

Those who wish to examine the derivations for Eqs. (2.20) and (2.21) will find them in Appendix B.

Before concluding this chapter, let's explore a simple application for the scalar triple product.

One popular method for detecting back-facing polygons is to compute the angle between the polygon's surface normal and the vector representing the observer's line of sight. This is shown in Fig. 2.28.

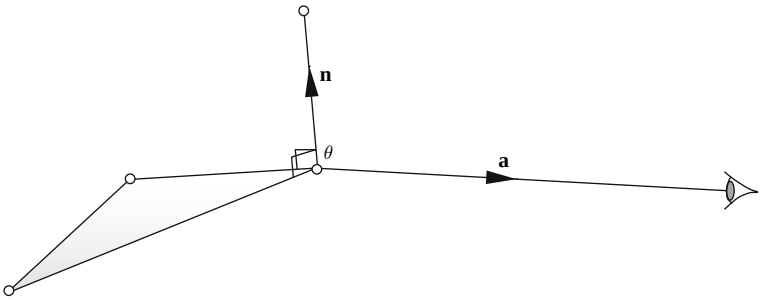


FIGURE 2.28.

If the angle θ between vectors **a** and **n** equals, or exceeds 90°, the polygon is invisible. But say such a normal was not available. Then the scalar triple product could be used instead.

First, we must know the order of the polygon's vertices as viewed from the outside. Let's assume that this is counter-clockwise. We then establish two vectors, **b** and **c**, such that their cross product points outwards. For example, in Fig. 2.29 the vertices A, B, and C would produce the vectors $\mathbf{b} = \overrightarrow{AB}$ and $\mathbf{c} = \overrightarrow{AC}$. We create a third vector **a** pointing towards the observer located at O, such that $\mathbf{a} = \overrightarrow{AO}$. This creates a tetrahedron whose volume can be positive, zero, or negative, depending on the orientation of the observer and polygon.

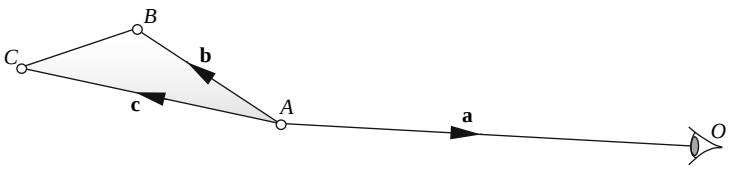


FIGURE 2.29.

The tetrahedron's volume is given by

$$\tfrac{1}{6}[\mathbf{abc}] = \tfrac{1}{6}\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \tfrac{1}{6} \begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix}.$$

If the volume is positive, the polygon is facing the observer; if the volume is zero, the polygon is parallel to the observer; and if the volume is negative, the polygon is facing away from the observer. Let's test this hypothesis with an example.

Figure 2.30 shows a scenario where the observer is located at (1, 0, 0) looking at one of the triangle's vertices positioned at the origin.

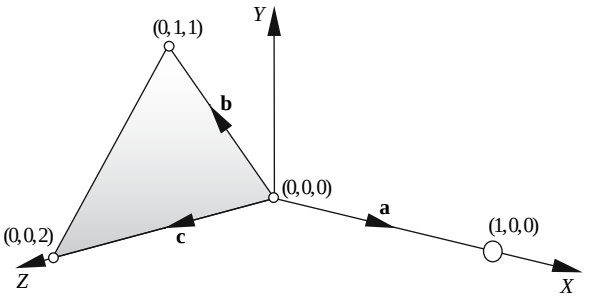


FIGURE 2.30.

Therefore, the vectors are $\mathbf{a} = \mathbf{i}, \quad \mathbf{b} = \mathbf{j} + \mathbf{k}, \quad \mathbf{c} = 2\mathbf{k}$, and the prism's volume is given by

$$Vol = \frac{1}{6}\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{vmatrix} = \frac{1}{6} \times 2$$

which means that the polygon is visible. If we locate the observer at (0, 0, −1) along the negative z-axis, the volume is zero:

$$Vol = \frac{1}{6}\begin{vmatrix} 0 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{vmatrix} = 0$$

which means that the polygon is invisible. Finally, if we locate the observer at (−1, 0, 0), the volume is negative:

$$Vol = \frac{1}{6}\begin{vmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{vmatrix} = \frac{1}{6} \times (-2)$$

which means that the polygon is invisible.

If the observer is located at the origin, there is no need to create any vectors at all as the volume is given by

$$Vol = \frac{1}{6}\begin{vmatrix} x_A & y_A & z_A \\ x_B & y_B & z_B \\ x_C & y_C & z_C \end{vmatrix}$$

where only the coordinates of the vertices are required. Unfortunately, this determinant provides a negative value for visible and a positive value for invisible.

2.11 Perpendicular vectors

When working in 2D, we often employ vectors that are perpendicular to some reference vector. For example, Fig. 2.31 shows two vectors **v** and **n**, where **n** is perpendicular to **v** and is expressed mathematically as $\mathbf{n} \perp \mathbf{v}$.

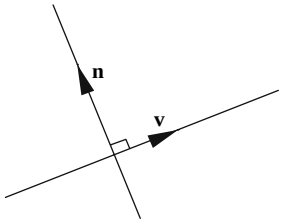


FIGURE 2.31.

The symbol '$\perp$' (pronounced "perp") can be regarded as an operator, such that given a vector **v**, $\mathbf{v}^{\perp}$ is a vector perpendicular to **v**. However, there is a problem: which way does $\mathbf{v}^{\perp}$ point? Figure 2.32 shows the two possibilities. Either way is valid. However, mathematics does employ a convention where a counter-clockwise rotation is positive, which also supports the right-handed axial system used in this book. Consequently, the orientation shown in Fig. 2.32(a) is the one adopted. The next step is to find the components of $\mathbf{v}^{\perp}$.

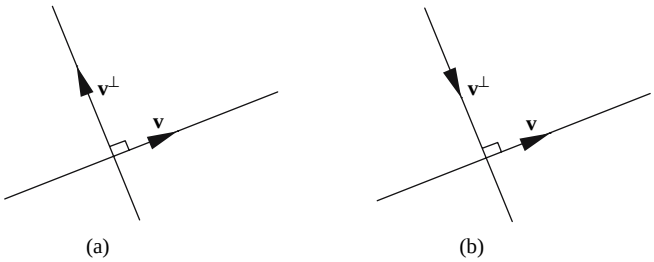


FIGURE 2.32.

Figure 2.33 shows vector $\mathbf{v} = a\mathbf{i} + b\mathbf{j}$, which makes an angle α with the x-axis, and $90° - \alpha$ with the y-axis. If we transpose **v**'s components to $-b\mathbf{i} + a\mathbf{j}$, we create a second vector, which must be perpendicular to **v**, because the angle between the two vectors is 90°. Therefore, we can state that

$$\begin{aligned} \mathbf{v} &= a\mathbf{i} + b\mathbf{j} \\ \mathbf{v}^{\perp} &= -b\mathbf{i} + a\mathbf{j} \end{aligned} \tag{2.22}$$

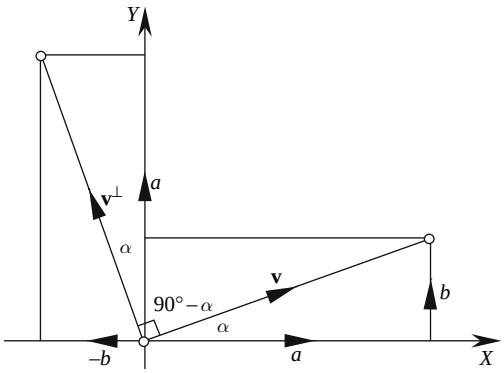


FIGURE 2.33.

Further confirmation is found in the dot product where

$$\mathbf{v} \cdot \mathbf{v}^{\perp} = (a\mathbf{i} + b\mathbf{j}) \cdot (-b\mathbf{i} + a\mathbf{j}) = -ab + ab = 0.$$

A second way of demonstrating this is to convert **v** into a complex number:
i.e.,

$$a\mathbf{i} + b\mathbf{j} \equiv a + ib \tag{2.23}$$

where

$$i = \sqrt{-1}.$$

If we multiply $a + ib$ by i, it effectively rotates it through 90°:

$$i\,(a + ib) = ai + i^2 b = -b + ai$$

where

$$-b + ia \equiv -b\mathbf{i} + a\mathbf{j}.$$

Although the sign change and component switching take a simple operation, it can be represented formally by this determinant:

$$\mathbf{v}^{\perp} = -\begin{vmatrix} \mathbf{i} & \mathbf{j} \\ a & b \end{vmatrix} = -b\mathbf{i} + a\mathbf{j}.$$

It may be obvious that the magnitude $\| \mathbf{v}^{\perp} \|$ equals the magnitude $\| \mathbf{v} \|$. If not, it can be shown that $\| \mathbf{v}^{\perp} \| = \sqrt{(-y_v)^2 + x_v^2} = \sqrt{x_v^2 + y_v^2} = \| \mathbf{v} \|$.

Now we already know that the dot product reveals the angle between two vectors:

$$\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \, \| \mathbf{b} \| \cos\theta$$

and

$$\cos\theta = \frac{\mathbf{a}\cdot\mathbf{b}}{\|\,\mathbf{a}\,\|\,\|\,\mathbf{b}\,\|}.$$

But what we don't know is the sense of θ; that is, is $\mathbf{b}$ rotated in a counter-clockwise direction relative to $\mathbf{a}$, or vice versa? Well, consider the scenario shown in Fig. 2.34, where $\mathbf{b}$ *is* rotated in a counter-clockwise direction relative to $\mathbf{a}$, creating a positive angle θ.

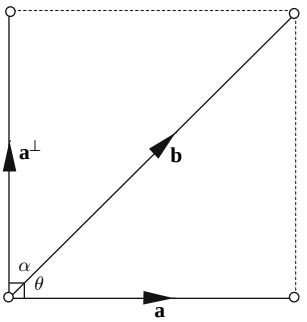


FIGURE 2.34.

From the figure we can state that

$$\mathbf{a}^{\perp}\cdot\mathbf{b} = \|\,\mathbf{a}^{\perp}\,\|\,\|\,\mathbf{b}\,\|\cos\alpha$$

and

$$\cos\alpha = \frac{\mathbf{a}^{\perp}\cdot\mathbf{b}}{\|\,\mathbf{a}^{\perp}\,\|\,\|\,\mathbf{b}\,\|} = \frac{\mathbf{a}^{\perp}\cdot\mathbf{b}}{\|\,\mathbf{a}\,\|\,\|\,\mathbf{b}\,\|}. \tag{2.24}$$

But note that as $\theta + \alpha = 90°$,

$$\sin\theta = \cos\alpha. \tag{2.25}$$

Substituting Eq. (2.25) in Eq. (2.24), we get

$$\sin\theta = \frac{\mathbf{a}^{\perp}\cdot\mathbf{b}}{\|\,\mathbf{a}\,\|\,\|\,\mathbf{b}\,\|}. \tag{2.26}$$

Now, the nature of the cosine function is that

$$\cos\theta = \cos(-\theta)$$

which is not the case for the sine function:

$$\sin(-\theta) = -\sin\theta.$$

This implies that Eq. (2.26) is sensitive to the direction of the angle θ. Thus, if $\mathbf{b}$ is rotated counterclockwise θ from $\mathbf{a}$, $\sin\theta$ is positive. But if $\mathbf{b}$ is rotated clockwise θ from $\mathbf{a}$, $\sin\theta$ is negative.

A simple example demonstrates the angular sensitivity of Eq. (2.26).
With reference to Fig. 2.34: let

$$\mathbf{a} = \mathbf{i}$$
$$\mathbf{b} = \mathbf{i} + \mathbf{j}$$
$$\mathbf{a}^{\perp} = \mathbf{j}$$

Therefore,

$$\sin\theta = \frac{\mathbf{a}^{\perp} \cdot \mathbf{b}}{\| \mathbf{a} \| \, \| \mathbf{b} \|} = \frac{\mathbf{j} \cdot (\mathbf{i} + \mathbf{j})}{1 \cdot \sqrt{2}} = \tfrac{1}{\sqrt{2}}.$$

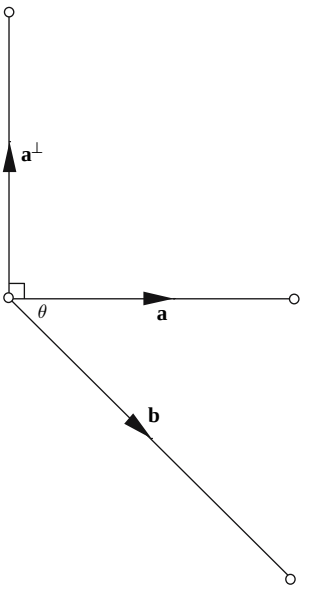


FIGURE 2.35.

In contrast, with reference to Fig. 2.35,

$$\mathbf{a} = \mathbf{i}$$
$$\mathbf{b} = \mathbf{i} - \mathbf{j}$$
$$\mathbf{a}^{\perp} = \mathbf{j}$$

Therefore,

$$\sin\theta = \frac{\mathbf{j} \cdot (\mathbf{i} - \mathbf{j})}{1 \cdot \sqrt{2}} = -\tfrac{1}{\sqrt{2}}.$$

Thus, we see that Eq. (2.26) is sensitive to the sign of the angle between the vectors.

In Section 2.12, when we discussed the vector product, we discovered that

$$\| \mathbf{a} \times \mathbf{b} \| = \| \mathbf{a} \| \, \| \mathbf{b} \| \sin \theta$$

which equals the area of the parallelogram formed by the vectors **a** and **b** (Fig. 2.22). But from Eq. (2.26) we can state that

$$\mathbf{a}^{\perp} \cdot \mathbf{b} = \| \mathbf{a} \| \, \| \mathbf{b} \| \sin \theta$$

Therefore, $\mathbf{a}^{\perp} \cdot \mathbf{b}$ equals the signed area of the parallelogram formed by **a** and **b**, as shown in Fig. 2.36.

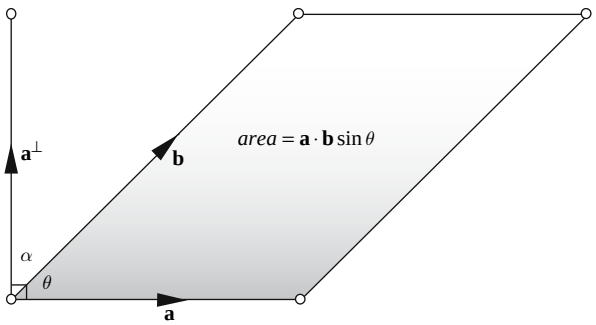


Figure 2.36.

2.12 Interpolating vectors

Interpolating between a pair of scalar quantities arises everywhere in computer graphics, especially in the control of light intensity, colour, position, or the orientation of objects and cameras. Vectors can also be interpolated and are regularly used in Phong shading, which is covered in Section 10.3. However, a consequence of linearly interpolated vectors is that vector length is not preserved, which can cause problems in the form of null vectors or the need for normalization. If vector length is important, then spherical interpolation should be used. In this section we consider both schemes.

2.12.1 Linear interpolation

Given two scalars v_1 and v_2, it is possible to linearly interpolate between them using

$$v = (1 - \lambda) v_1 + \lambda v_2$$

where

$$\lambda \text{ is a scalar such that } 0 \leqslant \lambda \leqslant 1.$$

When $\lambda < 0$ or $\lambda > 1$, interpolation continues, but $v < v_1$ and $v > v_2$, respectively.

It should not be a surprise that the same process also works with vectors, and we can write

$$\mathbf{w} = (1-\lambda)\,\mathbf{u} + \lambda\mathbf{v}, \quad 0 \leqslant \lambda \leqslant 1 \tag{2.27}$$

Figure 2.37 illustrates values of $\mathbf{w}$ for different values of λ.

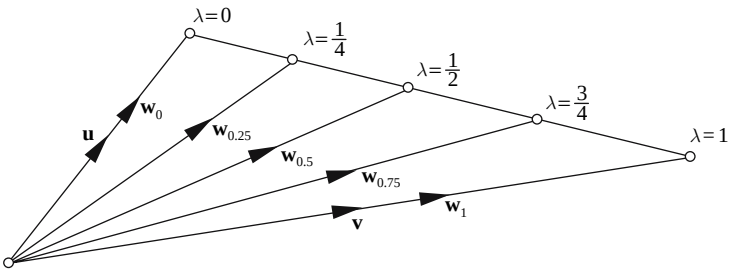


Figure 2.37.

In the following chapters we will employ Eq. (2.27) to identify points that lie on a line determined by two vectors.

2.12.2 Spherical interpolation

Spherical interpolation ensures that the magnitude of the interpolated vector is not comprised.

Figure 2.38 shows two unit vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ separated by an angle θ, which can be determined using

$$\theta = \cos^{-1}(\hat{\mathbf{u}} \cdot \hat{\mathbf{v}}).$$

From the figure, we see that

$$OT = a,\ TW = b$$

and TW is parallel with $\hat{\mathbf{v}}$.

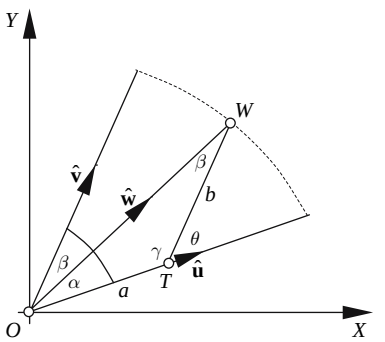


Figure 2.38.

The linear interpolation of the angle is controlled by the parameter λ, where

$$\alpha = \lambda\theta \text{ and } \beta = (1-\lambda)\,\theta$$

so that

$$\hat{\mathbf{w}} = a\hat{\mathbf{u}} + b\hat{\mathbf{v}}.$$

From Fig. 2.38 we get

$$\gamma = \pi - \theta$$

and

$$\sin\gamma = \sin\theta.$$

Therefore, using the sine rule:

$$\frac{OW}{\sin\gamma} = \frac{1}{\sin\theta} = \frac{a}{\sin\beta} = \frac{b}{\sin\alpha}$$

$$a = \frac{\sin\beta}{\sin\theta} = \frac{\sin(1-\lambda)\,\theta}{\sin\theta}$$

and

$$b = \frac{\sin\alpha}{\sin\theta} = \frac{\sin\lambda\theta}{\sin\theta}.$$

Finally,

$$\hat{\mathbf{w}} = \frac{\sin(1-\lambda)\,\theta}{\sin\theta}\hat{\mathbf{u}} + \frac{\sin\lambda\theta}{\sin\theta}\hat{\mathbf{v}} \tag{2.28}$$

Equation (2.28) is called a *slerp*, because in its 3D application the unit vectors are spherically interpolated across a unit-radius sphere. This interpolant can also be used for unit quaternions.

Let's test Eq. (2.28) with a simple example.

Given $\hat{\mathbf{u}} = \mathbf{i}$ and $\hat{\mathbf{v}} = \mathbf{j}$, then $\theta = 90°$.

Now let's find the interpolated vector when $\lambda = 0.5$.

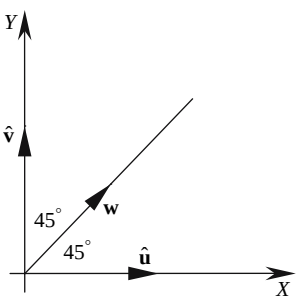


Figure 2.39.

As shown in Fig. 2.39, the interpolated vector should be

$$\mathbf{w} = \frac{\sqrt{2}}{2}\mathbf{i} + \frac{\sqrt{2}}{2}\mathbf{j}.$$

Using Eq. (2.31), we find that

$$\mathbf{w} = \frac{\sin 45^\circ}{\sin 90^\circ}\mathbf{i} + \frac{\sin 45^\circ}{\sin 90^\circ}\mathbf{j}$$
$$\mathbf{w} = \frac{\sqrt{2}}{2}\mathbf{i} + \frac{\sqrt{2}}{2}\mathbf{j}.$$

Although Eq. (2.31) assumes that $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are unit vectors, it will work with non-unit vectors. For example, if $\mathbf{u} = 2\mathbf{i}$ and $\mathbf{v} = 4\mathbf{j}$, then $\theta = 90^\circ$.

Now let's find the interpolated vector when $\lambda = 0.5$.

Therefore,

$$\mathbf{w} = \frac{\sqrt{2}}{2}2\mathbf{i} + \frac{\sqrt{2}}{2}4\mathbf{j} = \sqrt{2}\mathbf{i} + 2\sqrt{2}\mathbf{j}.$$

The magnitude of $\mathbf{w}$ is

$$\| \mathbf{w} \| = \sqrt{\left(\sqrt{2}\right)^2 + \left(2\sqrt{2}\right)^2} = \sqrt{10} = 3.162.$$

So, even though the rotated angle is 50% of the separating angle, the vector's length exceeds the half-way point between 2 and 4, which is because the interpolation is spherical rather than linear.

If the length of the interpolated vector is required to be linearly related to θ, then we must normalize $\mathbf{u}$ and $\mathbf{v}$ and scale the interpolated vector as follows:

$$\mathbf{w} = [(1-\lambda)\,\| \mathbf{u} \| + \lambda\,\| \mathbf{v} \|] \left(\frac{\sin(1-\lambda)\,\theta}{\sin\theta}\hat{\mathbf{u}} + \frac{\sin\lambda\theta}{\sin\theta}\hat{\mathbf{v}} \right).$$

For instance, using the above example, where $\mathbf{u} = 2\mathbf{i}$, $\mathbf{v} = 4\mathbf{j}$, and $\lambda = 0.5$

then

$$\mathbf{w} = (0.5 \times 2 + 0.5 \times 4) \left(\frac{\sin 45^\circ}{\sin 90^\circ}\mathbf{i} + \frac{\sin 45^\circ}{\sin 90^\circ}\mathbf{j} \right)$$

and

$$\mathbf{w} = 3 \left(\frac{\sqrt{2}}{2}\mathbf{i} + \frac{\sqrt{2}}{2}\mathbf{j} \right)$$

which are correct.

Needless to say, the interpolation works in 3D as well as 2D.

2.13 Direction cosines

A vector's orientation and magnitude are encoded within its Cartesian components. The magnitude of a 3D vector **v** is readily obtained using

$$\| \mathbf{v} \| = \sqrt{x_v^2 + y_v^2 + z_v^2}$$

while its orientation is obtained using

$$\alpha = \tan^{-1}\left(\frac{y_v}{x_v}\right)$$

where α is the angle between the vector and the x-axis. However, we can develop this idea of orientation using *direction cosines*.

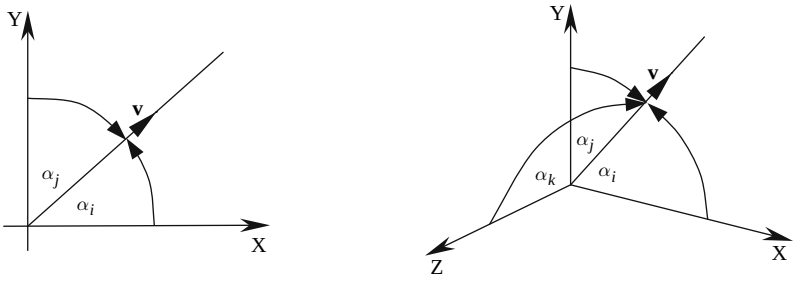


FIGURE 2.40.

Figure 2.40 illustrates the idea behind direction cosines for a 2D and a 3D vector. The angles α_i, α_j, and α_k (for a 3D vector) provide an elegant visual mechanism for orientating a vector relative to the Cartesian axes. In fact, the cosines of the angles are used as they are intimately related to the vector's components.

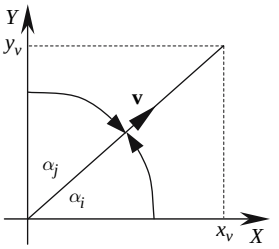


FIGURE 2.41.

To begin with, consider the 2D vector **v** in Fig. 2.41. It has components x_v and y_v and it is obvious that $\alpha_i + \alpha_j = 90°$. Now

$$\cos\alpha_i = \frac{x_v}{\| \mathbf{v} \|}$$

and

$$\cos\alpha_j = \frac{y_v}{\|\mathbf{v}\|}.$$

If $\mathbf{v}$ is a unit vector,

$$\cos\alpha_i = v_x$$

and

$$\cos\alpha_j = v_y$$

and are called the direction cosines of $\mathbf{v}$.

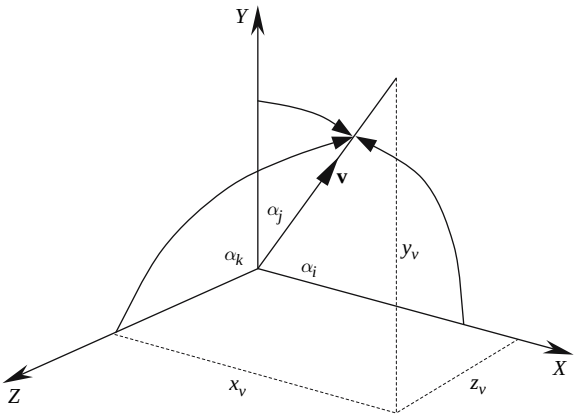


FIGURE 2.42.

In the 3D scenario shown in Fig. 2.42, it may not be immediately obvious what $\cos\alpha_i$, $\cos\alpha_j$, and $\cos\alpha_k$ are, but the dot product reveals the answer.

To find $\cos\alpha_i$,

$$\mathbf{v}\cdot\mathbf{i} = \|\mathbf{v}\|\,\|\mathbf{i}\|\cos\alpha_i = x_v$$

Therefore,

$$\cos\alpha_i = \frac{x_v}{\|\mathbf{v}\|}.$$

Similarly,

$$\cos\alpha_j = \frac{y_v}{\|\mathbf{v}\|}$$

and

$$\cos\alpha_k = \frac{z_v}{\|\mathbf{v}\|}.$$

Once more, if $\mathbf{v}$ is a unit vector, the direction cosines are the vector's components.

2.14 Change of axial system

One useful application for direction cosines is in the transformation of points from one coordinate system to another. For example, consider the problem of calculating the coordinates of a 2D point relative to a system of axes that have been rotated θ relative to the original axes, as shown in Fig. 2.43.

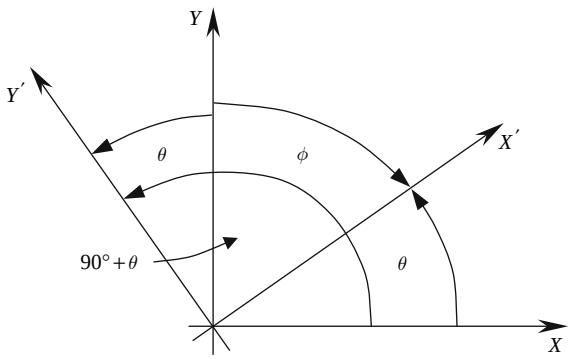


FIGURE 2.43.

X' is rotated θ relative to X and has direction cosines $\cos\theta$ and $\cos\phi$.

But as

$$\sin\theta = \cos\phi$$

X' has direction cosines $\cos\theta$ and $\sin\theta$.

Similarly, Y' is rotated relative to Y and has direction cosines $\cos(90° + \theta)$ and $\cos\theta$.

But as

$$\cos(90° + \theta) = -\sin\theta$$

Y' has direction cosines $-\sin\theta$ and $\cos\theta$.

If the axial system is rotated θ, this is equivalent to rotating a point in the opposite direction $-\theta$.

The rotation transform for an angle θ is

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Therefore, the transform for an angle $-\theta$ is

$$\begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix}$$

which becomes

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

But notice that the top-row elements of the transform are the direction cosines of the rotated X'-axis, and the bottom-row elements are the direction cosines of the rotated Y'-axis. Thus, we can state that the coordinates of a point $P(x_P, y_P)$ in a rotated frame of reference are given by

$$\begin{bmatrix} x'_P \\ y'_P \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \cdot \begin{bmatrix} x_P \\ y_P \end{bmatrix}$$

where

r_{11} is the direction cosine of X' relative to X

r_{12} is the direction cosine of X' relative to Y

r_{21} is the direction cosine of Y' relative to X

r_{22} is the direction cosine of Y' relative to Y.

Note that because $\sin^2\theta + \cos^2\theta = 1$, the direction cosines for any vector possess the following qualities:

$$r_{11}^2 + r_{12}^2 = 1$$
$$r_{21}^2 + r_{22}^2 = 1.$$

To demonstrate this transform in action, consider the case of calculating the new coordinates of $P(1, 1)$, where the axes are rotated 45°. This scenario is shown in Fig. 2.44.

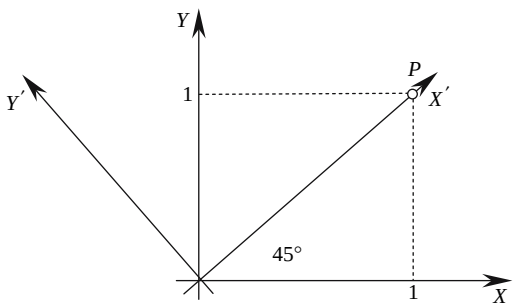


Figure 2.44.

It should be obvious from Fig. 2.44 that the coordinates of P relative to the rotated axes are $\left(\sqrt{2}, 0\right)$, and the transform will confirm this result.

The direction cosines for X' are $\cos 45°$ and $\sin 45°$, while the direction cosines for Y' are $-\sin 45°$ and $\cos 45°$, which produces the following transform:

$$\begin{bmatrix} x'_P \\ y'_P \end{bmatrix} = \begin{bmatrix} 0.707 & 0.707 \\ -0.707 & 0.707 \end{bmatrix} \cdot \begin{bmatrix} x_P \\ y_P \end{bmatrix}$$
$$(x'_P,\ y'_P) = (1.414,\ 0)\,.$$

A similar argument can be used for calculating the new coordinates of a point relative to a rotated set of 3D axes. Once again, the point is rotated in the opposite direction to the axis rotation. Although this can require three individual rotations about the x-, y-, and z-axes, the final result can be expressed as a transform:

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

where

$r_{11}\, r_{12}\, r_{13}$ are the direction cosines of the rotated X'-axis relative to the original X-, Y-, and Z-axes, respectively,
$r_{21}\, r_{22}\, r_{23}$ are the direction cosines of the rotated Y'-axis relative to the original X-, Y-, and Z-axes, respectively,
$r_{31}\, r_{32}\, r_{33}$ are the direction cosines of the rotated Z'-axis relative to the original X-, Y-, and Z-axes, respectively.

Thus, the new coordinates of a point $P\,(x_P, y_P, z_P)$ are given by

$$\begin{bmatrix} x'_P \\ y'_P \\ z'_P \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix}.$$

Note that

$$\begin{aligned} r_{11}^2 + r_{12}^2 + r_{13}^2 &= 1 \\ r_{21}^2 + r_{22}^2 + r_{23}^2 &= 1 \\ r_{31}^2 + r_{32}^2 + r_{33}^2 &= 1. \end{aligned}$$

To demonstrate the above, consider the case of calculating the new coordinates of $P\,(1, 1, 1)$, where the axial system is rotated as illustrated in Fig. 2.45.

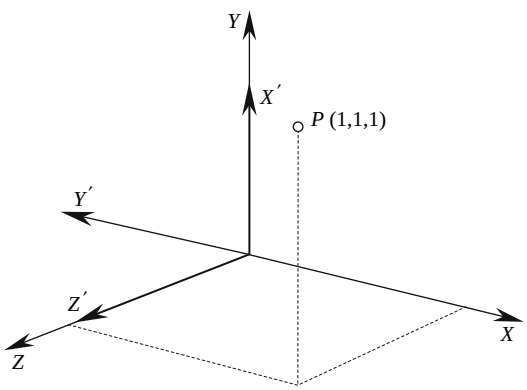


FIGURE 2.45.

From Fig. 2.45, we can predict that the new coordinates are $(1, -1, 1)$.

The direction cosines for X', Y', and Z' are $(0, 1, 0)$, $(-1, 0, 0)$, $(0, 0, 1)$, respectively.

Therefore,

$$\begin{bmatrix} x'_P \\ y'_P \\ z'_P \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

which is correct!

2.15 Summary

- A vector quantity possesses both magnitude and direction, which are encoded within its components.
- Given two points A and B, the vector $\overrightarrow{AB}$ is given by

$$\overrightarrow{AB} = \begin{bmatrix} x_B - x_A \\ y_B - y_A \\ z_B - z_A \end{bmatrix}$$

- The magnitude or length of a vector is expressed as $\| \mathbf{a} \|$, which equals

$$\sqrt{x_a^2 + y_a^2} \text{ or } \sqrt{x_a^2 + y_a^2 + z_a^2}.$$

- The resultant of two vectors $\mathbf{a}$ and $\mathbf{b}$ equals their vector sum $\mathbf{a} + \mathbf{b}$.
- Vector addition obeys the commutative and associative laws:

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} \qquad \text{(commutative law)}$$
$$\mathbf{a} + (\mathbf{b} + \mathbf{c}) = (\mathbf{a} + \mathbf{b}) + \mathbf{c} \qquad \text{(associative law)}.$$

- The scalar product of two vectors is defined as

$$\mathbf{a} \cdot \mathbf{b} = \| \mathbf{a} \| \, \| \mathbf{b} \| \cos\theta = x_a x_b + y_a y_b + z_a z_b$$

  where θ is the angle between the vectors.
- The scalar product obeys the commutative and distributive laws of multiplication:

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$$
$$\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}.$$

- The vector product of two vectors is defined as

$$\mathbf{a} \times \mathbf{b} = \mathbf{c} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_a & y_a & z_a \\ x_b & y_b & z_b \end{vmatrix}$$

  where $\| \mathbf{c} \| = \| \mathbf{a} \| \, \| \mathbf{b} \| \sin\theta$ and θ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

- If P and Q, with position vectors $\mathbf{p}$ and $\mathbf{q}$, respectively, are two points on a line perpendicular to a vector $\mathbf{v}$, then $\mathbf{p} \cdot \mathbf{v} = \mathbf{q} \cdot \mathbf{v}$.
- The vector product is non-commutative:

$$\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$$

  but obeys the distributive law:

$$\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}.$$

- The triple product or box product is given by

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix}.$$

- The volume of a prism constructed from three vectors is given by

$$Volume = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

- The vector triple product is given by

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}$$

  and

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\,\mathbf{a}.$$

- The perpendicular to a vector $\mathbf{v} = a\mathbf{i} + b\mathbf{j}$ is $\mathbf{v}^{\perp} = -b\mathbf{i} + a\mathbf{j}$.
- The magnitude of $\mathbf{v}^{\perp}$ equals the magnitude of $\mathbf{v}$.
- The perp product equals

$$\mathbf{a}^{\perp} \cdot \mathbf{b} = \| \mathbf{a} \| \, \| \mathbf{b} \| \sin\theta$$

  which equals the signed area of the parallelogram formed by $\mathbf{a}$ and $\mathbf{b}$.
- Two vectors can be linearly interpolated using

$$\mathbf{w} = (1 - \lambda)\,\mathbf{u} + \lambda \mathbf{v} \qquad 0 \leq \lambda \leq 1$$

- Two unit vectors can be spherically interpolated using

$$\hat{\mathbf{w}} = \frac{\sin(1-\lambda)\,\theta}{\sin\theta}\hat{\mathbf{u}} + \frac{\sin\lambda\theta}{\sin\theta}\hat{\mathbf{v}} \qquad 0 \le \lambda \le 1$$

  where θ is the angle between $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$.
- The direction cosines of a vector are the cosines of the angles between the vector and the Cartesian axes.
- A point $P(x_P, y_P)$ is rotated to $P'(x'_P, y'_P)$ using

$$\begin{bmatrix} x'_P \\ y'_P \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \cdot \begin{bmatrix} x_P \\ y_P \end{bmatrix}$$

  where

  r_{11} r_{12} are the direction cosines of the x-axis, and
  r_{21} r_{22} are the direction cosine of the y-axis.
  $r_{11}^2 + r_{12}^2 = 1$ and $r_{21}^2 + r_{22}^2 = 1$.

- A point $P(x_P, y_P, z_P)$ is rotated to $P'(x'_P, y'_P, z'_P)$ using

$$\begin{bmatrix} x'_P \\ y'_P \\ z'_P \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} x_P \\ y_P \\ z_P \end{bmatrix}$$

  where

  r_{11} r_{12} r_{13} are the direction cosines of the x-axis,
  r_{21} r_{22} r_{23} are the direction cosines of the y-axis, and
  r_{31} r_{32} r_{33} are the direction cosines of the y-axis.
  $r_{11}^2 + r_{12}^2 + r_{13}^2 = 1$, $r_{21}^2 + r_{22}^2 + r_{23}^2 = 1$, $r_{31}^2 + r_{32}^2 + r_{33}^2 = 1$.

3 Straight Lines

3.1 Introduction

Now that we have explored the basics of vector analysis, let us apply some of these ideas to the description of straight lines. There are various ways of describing straight lines; these include

the normal form of the line equation $y = mx + c$,
the general form of the line equation $ax + by + c = 0$,
the Cartesian form of the line equation $ax + by = c$,
and the parametric form of the line equation $\mathbf{p} = \mathbf{t} + \lambda\hat{\mathbf{v}}$.

The first three only apply to 2D straight lines, whereas the fourth definition applies to 2D and 3D straight lines. So let us begin with this latter equation.

3.2 The parametric form of the line equation

The first objective is to find a way to move along a straight line using a scalar parameter. Fortunately, vectors provide such a mechanism in the form of the operation $\lambda\hat{\mathbf{v}}$, which scales the unit vector $\hat{\mathbf{v}}$ with the scalar λ. But we also need to relate our position relative to some reference point, which must be included in our definition. We begin as follows.

Figure 3.1 shows a line that passes through the points T and P and is parallel with the vector $\hat{\mathbf{v}}$. From this scenario we can state

$$\overrightarrow{OP} = \overrightarrow{OT} + \overrightarrow{TP}.$$

Let

$\hat{\mathbf{v}}$ = the unit vector defining the line's direction,
λ = a scalar that acts as a parameter to position us anywhere along the direction of $\hat{\mathbf{v}}$,
T = a reference point on the line,

$\mathbf{t}$ = the position vector for T, and
$\mathbf{p}$ = the position vector for any point $P(x, y)$ on the line.

Then

$$\mathbf{p} = \mathbf{t} + \lambda\hat{\mathbf{v}}. \tag{3.1}$$

It should be obvious that Eq. (3.1) works equally well in R^3 as it does in R^2, so let us bring it alive using the coordinates used in Fig. 3.1.

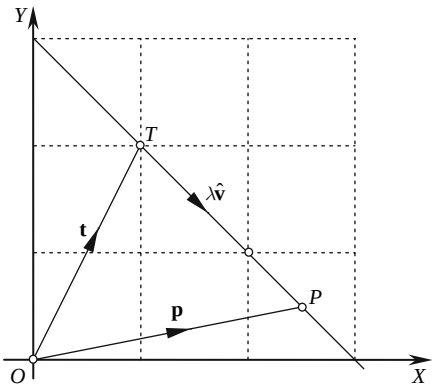


FIGURE 3.1.

We see that

$$\mathbf{t} \text{ points to } T(1, 2)$$

and

$$\mathbf{v} = \mathbf{i} - \mathbf{j}$$

Therefore,

$$\hat{\mathbf{v}} = \tfrac{1}{\sqrt{2}}(\mathbf{i} - \mathbf{j}).$$

When

$$\lambda = 0$$

$$x = 1 \text{ and } y = 2$$

and when

$$\lambda = 1$$

$$x = 1 + \tfrac{1}{\sqrt{2}} = 1.7071$$

$$y = 2 - \tfrac{1}{\sqrt{2}} = 1.2929.$$

Thus, (1.707, 1.293) is a point one unit along $\hat{\mathbf{v}}$ from T. By changing the value and sign of λ, we can move forward or backward along $\hat{\mathbf{v}}$.

Let's illustrate Eq. (3.1) in a 3D context, as shown in Fig. 3.2.
Let

$$\mathbf{t} \text{ point to } T\,(1, 2, 0)$$

and

$$\mathbf{v} = -\mathbf{i} - 2\mathbf{j} + 2\mathbf{k}.$$

Therefore,

$$\hat{\mathbf{v}} = \tfrac{1}{3}\,(-\mathbf{i} - 2\mathbf{j} + 2\mathbf{k})\,.$$

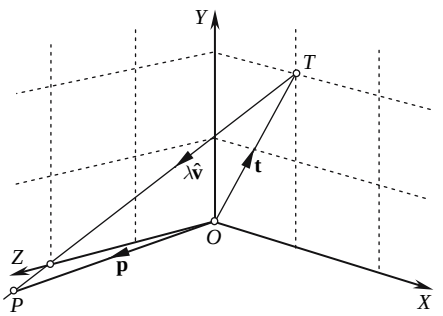


Figure 3.2.

The vector equation is

$$\mathbf{p} = \mathbf{t} + \lambda\hat{\mathbf{v}}.$$

When

$$\lambda = 0$$
$$x = 1, \quad y = 2 \text{ and } z = 0$$

and when

$$\lambda = 1$$
$$x = 1 - \tfrac{1}{3} = \tfrac{2}{3}$$
$$y = 2 - \tfrac{2}{3} = 1\tfrac{1}{3}$$
$$z = 0 + \tfrac{2}{3} = \tfrac{2}{3}.$$

Thus, $\left(\frac{2}{3}, 1\frac{1}{3}, \frac{2}{3}\right)$ is a point one unit along $\hat{\mathbf{v}}$ from T. Similarly, by changing the value and sign of λ, we can move forward and backward along the 3D line.

As we may not always have an explicit description of $\hat{\mathbf{v}}$, let's consider the case when the line is represented by two points in space. If the two points are $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$, we can state that

$$\mathbf{v} = (x_2 - x_1)\,\mathbf{i} + (y_2 - y_1)\,\mathbf{j}$$

from which we can compute $\| \mathbf{v} \|$.

$\hat{\mathbf{v}}$ is then equal to $\frac{\mathbf{v}}{\| \mathbf{v} \|}$ and T can be P_1.

We can now write Eq. (3.1) as follows:

$$\mathbf{p} = \mathbf{p}_1 + \lambda \hat{\mathbf{v}} \tag{3.2}$$

where $\mathbf{p}_1$ is the position vector for P_1. This scenario is shown in Fig. 3.3.

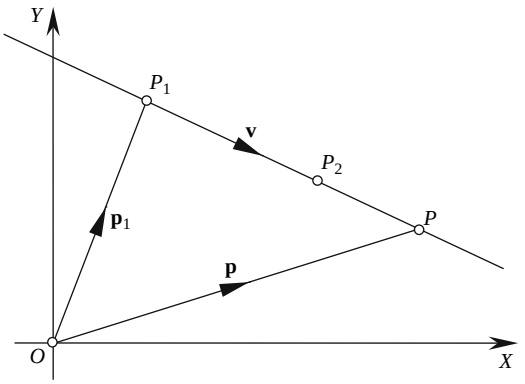


FIGURE 3.3.

In the current R^2 context, Eq. (3.2) reveals

$$x = x_1 + \lambda \hat{\mathbf{v}}$$
$$y = y_1 + \lambda \hat{\mathbf{v}}$$

which expands to

$$x = x_1 + \frac{\lambda}{\| \mathbf{v} \|}(x_2 - x_1)$$
$$y = y_1 + \frac{\lambda}{\| \mathbf{v} \|}(y_2 - y_1)$$

and

$$x = \left(1 - \frac{\lambda}{\| \mathbf{v} \|}\right) x_1 + \frac{\lambda}{\| \mathbf{v} \|} x_2 \tag{3.3}$$

$$y = \left(1 - \frac{\lambda}{\| \mathbf{v} \|}\right) y_1 + \frac{\lambda}{\| \mathbf{v} \|} y_2. \tag{3.4}$$

It is obvious from Eqs. (3.3) and (3.4) that when $\lambda = \| \mathbf{v} \|$, $x = x_2$ and $y = y_2$, and represents a point (x,y) $\| \mathbf{v} \|$ units along the line from $P_1 \cdot (x, y)$.

Let us follow this parametric description of a straight line with the derivation of the Cartesian form of the line equation.

3.3 The Cartesian form of the line equation

We already know that the Cartesian form of the line equation is $ax + by = c$, but what is not obvious is the geometric significance of a, b, and c. Fortunately, vector analysis provides an insight into these terms, which is the objective of this section.

We begin by setting out to discover the coordinates of any point $P(x, y)$ on a straight line. The strategy used may not immediately appear intuitive, but try following it, and hopefully the approach will become obvious.

Step 1

Construct a line away from the origin O containing the point $P(x, y)$ and its associated position vector $\mathbf{p}$, as shown in Fig. 3.4.

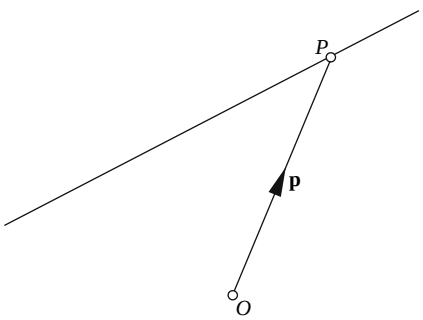


FIGURE 3.4.

We now need to construct a second path from O to P.

Step 2

Construct a line from O to a point C on the line such that $\overrightarrow{OC}$ is perpendicular to the line. It is convenient to make $\overrightarrow{OC}$ a scalar multiple of a unit vector $\hat{\mathbf{n}} = a\mathbf{i} + b\mathbf{j}$. i.e., $\overrightarrow{OC} = c\hat{\mathbf{n}}$, where c is some scalar as shown in Fig. 3.5.

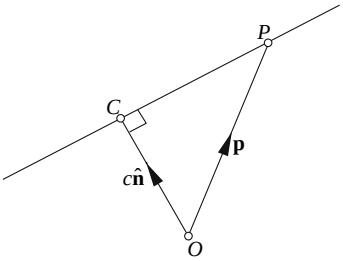


FIGURE 3.5.

Step 3

Extend the diagram to include the vector $\overrightarrow{CP}$, which completes the second path from O to P, as shown in Fig. 3.6.

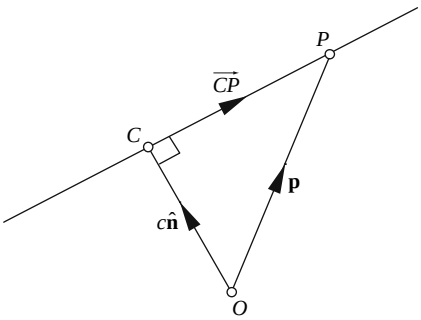


FIGURE 3.6.

We can now write $\overrightarrow{OP} = \overrightarrow{OC} + \overrightarrow{CP}$. Substituting vector names, we find that

$$\mathbf{p} = c\hat{\mathbf{n}} + \overrightarrow{CP}. \tag{3.5}$$

Although we know c and $\hat{\mathbf{n}}$ as well as the fact that $\mathbf{p}$ points to any point on the line, we do not know $\overrightarrow{CP}$. Somehow, $\overrightarrow{CP}$ has to be eliminated, which can be achieved by the following cunning subterfuge.

Remembering that the dot product of two perpendicular vectors is zero, $c\hat{\mathbf{n}} \cdot \overrightarrow{CP}$ must equal zero. In fact, the scalar c is superfluous; therefore, $\hat{\mathbf{n}} \cdot \overrightarrow{CP} = 0$.

Unfortunately, Eq. (3.5) does not contain such a term, but there is nothing to stop us from introducing one by multiplying Eq. (3.5) by $\hat{\mathbf{n}}$ using the dot product:

$$\hat{\mathbf{n}} \cdot \mathbf{p} = c\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} + \hat{\mathbf{n}} \cdot \overrightarrow{CP}$$

which now reduces to

$$\hat{\mathbf{n}} \cdot \mathbf{p} = c\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} \tag{3.6}$$

We are almost there, and something is staring us in the face, which we must learn to recognise: $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}$. As the angle between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}$ is zero degrees,

$$\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = \| \hat{\mathbf{n}} \| \, \| \hat{\mathbf{n}} \| = 1$$

Therefore, we can write Eq. (3.6) as

$$\hat{\mathbf{n}} \cdot \mathbf{p} = c. \tag{3.7}$$

Expanding Eq. (3.7), we get

$$(a\mathbf{i} + b\mathbf{j}) \cdot (x\mathbf{i} + y\mathbf{j}) = c$$
$$ax + by = c.$$

which we recognise as the Cartesian form of the line equation where

x, y are the coordinates of a point on the line,
a and b are the components of a unit vector normal to the line, and
c is the perpendicular distance from the origin to the line.

Although the above interpretation is correct for our example, we must be careful not to be too hasty when interpreting straight-line equations. For example, in the following equation,

$$3x + 4y = 10 \tag{3.8}$$

10 does not represent the perpendicular distance from the origin to the line. This is because the 3 and 4 are not the components of a unit vector; they are the components of a vector 5 units long.

Therefore, if we scale Eq. (3.8) by $\frac{1}{5}$, we obtain

$$0.6x + 0.8y = 2 \tag{3.9}$$

where 0.6 and 0.8 are the components of a unit vector, and 2 is the perpendicular distance from the origin to the line. Figure 3.7 shows the line equation representing Eq. (3.9).

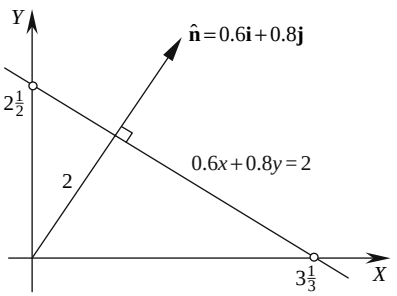


Figure 3.7.

So far, all this seems reasonable and consistent, but how do we interpret the following equation, where the c term is negative?

$$0.6x + 0.8y = -2 \tag{3.10}$$

Figure 3.8 shows what its graph looks like.

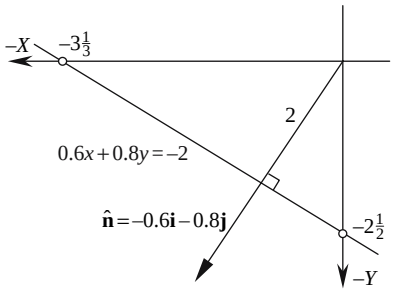


FIGURE 3.8.

Obviously, the Euclidean perpendicular distance from the origin to the line cannot be negative. Furthermore, Fig. 3.8 shows that the normal vector $\hat{\mathbf{n}} = -0.6\mathbf{i} - 0.8\mathbf{j}$, which provides the real clue. These two facts are simply reminding us to rewrite Eq. (3.10) in a form that preserves a positive c term, which is achieved by multiplying Eq. (3.10) by -1:

$$-0.6x - 0.8y = 2.$$

Now the vector components are correct, and the perpendicular distance from the origin to the line is $+2$. Nevertheless, Eq. (3.10) is still valid algebraically, it's just that a positive c term makes geometric sense and makes the normal vector components explicit.

3.4 The general form of the line equation

The general form of the line equation is

$$ax + by + c = 0.$$

The meanings of a, b, and c are still the same as before, all that we have done is to form a single expression that equals zero for any point on the line. Thus, Eq. (3.9) would be written as

$$0.6x + 0.8y - 2 = 0.$$

This is a convenient notation that lends itself to space partitioning. But before we look at this, consider how the following general form of the line equation is converted into its Cartesian equivalent:

$$0.6x + 0.8y + 2 = 0.$$

We begin by moving the c term to the right-hand side:

$$0.6x + 0.8y = -2.$$

Notice that the right-hand side is negative, which is compensated for by multiplying the whole equation by -1 to create

$$-0.6x - 0.8y = 2$$

which now makes geometric sense. Now let's examine how the general form of the line equation can be used to partition space.

3.5 2D space partitioning

One recurring problem in computer graphics is the need to know if a point is inside or outside a 2D area or 3D volume. There are various solutions to this problem, but let's see how vector analysis provides some assistance.

Equation (3.9) shows the Cartesian form of a line equation, whose graph is shown in Fig. 3.7. Rewriting Eq. (3.9) in its general form, we get

$$0.6x + 0.8y - 2 = 0 \tag{3.11}$$

When we substitute values of (x, y) for points on the line, the expression $0.6x + 0.8y - 2$ equals zero. But what happens when we substitute points that are on either side of the line? Well, let's try substituting some points.

The point $(10, 10)$ is obviously above the line, and when it is substituted in $0.6x + 0.8y - 2$, we obtain the value $+12$. On the other hand, the point $(-10, -10)$ is below the line, and when substituted in $0.6x + 0.8y - 2$ produces the value-16.

If we continued substituting other points, we would discover that the line partitions space into two zones: a zone above the line where $0.6x + 0.8y - 2$ is positive, and a zone below the line where $0.6x + 0.8y - 2$ is negative. But the terms "above" and "below" are incorrect, as they have no meaning for vertical lines!

To discover the key to this problem, we need to see what happens when the line's normal vector is reversed:

$$-0.6x - 0.8y - 2 = 0 \tag{3.12}$$

whose graph is shown in Fig. 3.8. Substituting $(10, 10)$ in $-0.6x - 0.8y - 2$ equals -16, whereas $(-10, -10)$ equals $+12$. Notice that for both equations, the positive sign is in the zone containing the normal vector, which is always the case. The reader may wish to reason why this is so.

Therefore, if we had a convex boundary formed by straight edges, how can we arrange for their normal vectors to point outwards or inwards? This is what we'll investigate next. Ideally, we want to create one of the scenarios shown in Fig. 3.9.

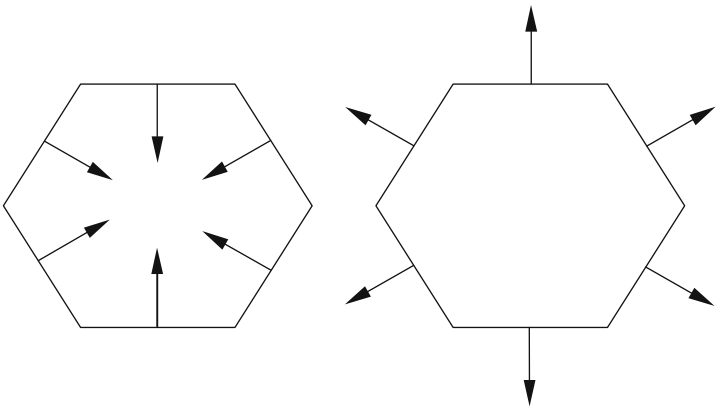


FIGURE 3.9.

We begin by making the convex boundary from a list of vertices P_1, P_2, P_3, and P_4, defined in a counter-clockwise sequence, which creates a corresponding chain of straight edges. As an example, a quadrilateral has been chosen as shown in Fig. 3.10. We now need to find the general form of the line equation for each boundary edge. If we do not apply a consistent algorithm, we could end up with an incoherent collection of normal vectors.

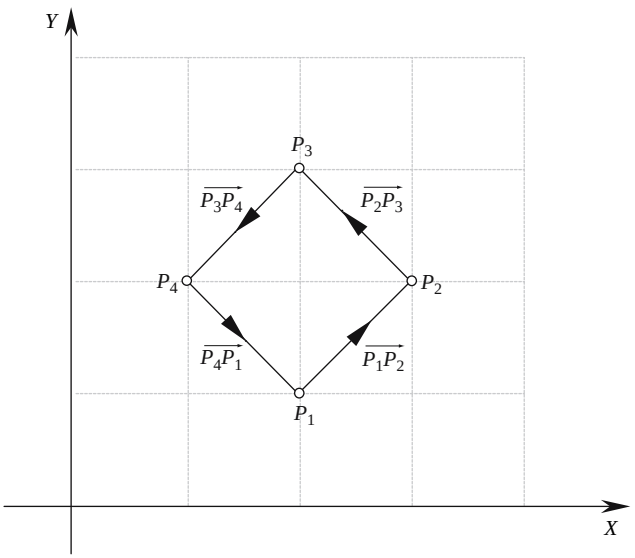


FIGURE 3.10.

We start with two points P_1 and P_2 that are the ends of an edge, as shown in Fig. 3.11.

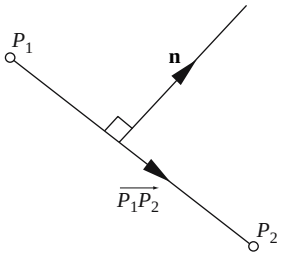


FIGURE 3.11.

Now we let

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} \text{ be perpendicular to } \overrightarrow{P_1P_2}$$

Therefore,

$$\overrightarrow{P_1P_2} \cdot \mathbf{n} = 0$$

and

$$\begin{gathered}((x_2 - x_1)\,\mathbf{i} + (y_2 - y_1)\,\mathbf{j}) \cdot (a\mathbf{i} + b\mathbf{j}) = 0 \\ a\,(x_2 - x_1) + b\,(y_2 - y_1) = 0 \\ ax_2 - ax_1 + by_2 - by_1 = 0 \\ \frac{a}{b} = \frac{y_1 - y_2}{x_2 - x_1}\end{gathered}$$

Therefore,

$$a = y_1 - y_2$$

and

$$b = x_2 - x_1.$$

The general form of the line equation is

$$(y_1 - y_2)\,x + (x_2 - x_1)\,y + c = 0 \tag{3.13}$$

c is found by substituting P_1 or P_2 in Eq. (3.13).

We now evaluate Eq. (3.13) for each boundary edge.

$\overrightarrow{P_1P_2}$

$$\begin{array}{ll} x_1 = 2 & y_1 = 1 \\ x_2 = 3 & y_2 = 2 \end{array}$$

Substituting these in Eq. (3.13), we get

$$-x+y+c=0 \tag{3.14}$$

Substituting the point $(2, 1)$ in Eq. (3.14) gives

$$-2+1+c=0 \qquad \therefore c=1$$

and the line equation is

$$-x+y+1=0$$

with a normal vector

$$-\mathbf{i}+\mathbf{j}.$$

$\overrightarrow{P_2P_3}$

$$\begin{aligned} x_1&=3 \quad y_1=2\\ x_2&=2 \quad y_2=3 \end{aligned}$$

Substituting these in Eq. (3.13), we get

$$-x-y+c=0 \tag{3.15}$$

Substituting the point $(3, 2)$ in Eq. (3.15) gives

$$-3-2+c=0 \qquad \therefore c=5$$

and the line equation is

$$-x-y+5=0$$

with a normal vector

$$-\mathbf{i}-\mathbf{j}.$$

$\overrightarrow{P_3P_4}$

$$\begin{aligned} x_1&=2 \quad y_1=3\\ x_2&=1 \quad y_2=2 \end{aligned}$$

Substituting these in Eq. (3.13), we get

$$x-y+c=0 \tag{3.16}$$

Substituting the point $(2, 3)$ in Eq. (3.16) gives

$$2-3+c=0 \qquad \therefore c=1$$

and the line equation is

$$x-y+1=0$$

with a normal vector

$$\mathbf{i}-\mathbf{j}.$$

$\overrightarrow{P_4P_1}$

$$x_1 = 1 \quad y_1 = 2$$
$$x_2 = 2 \quad y_2 = 1$$

Substituting these in Eq. (3.13), we get

$$x + y + c = 0 \tag{3.17}$$

Substituting the point $(1, 2)$ in Eq. (3.17) gives

$$1 + 2 + c = 0 \qquad \therefore c = -3$$

and the line equation is

$$x + y - 3 = 0$$

with a normal vector

$$\mathbf{i} + \mathbf{j}.$$

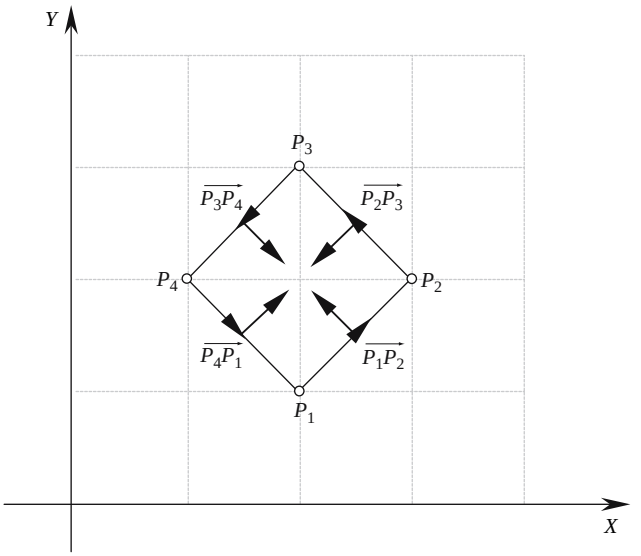


FIGURE 3.12.

Figure 3.12 shows the boundary with the normal vectors superimposed, and we see that all the normal vectors point to the boundary's inside. The four expressions can now be tested with different points, and it is the sign of the expression that confirms their space partition.

If only one expression is zero – the point is on the boundary.
If two expressions are zero – the point is a boundary vertex.
If any expression is negative – the point is outside the boundary.
If all expressions are positive – the point is inside the boundary.

Table 3.1 summarises some of these conditions.

TABLE 3.1

Point	$-x+y+1$	$-x-y+5$	$x-y+1$	$x+y-3$	Partition
(2, 2)	+	+	+	+	inside
(1, 1)	+	+	+	−	outside
(2, 0)	−	+	+	−	outside
(3, 1)	−	+	+	+	outside
(4, 2)	−	−	+	+	outside
(3, 3)	+	−	+	+	outside
(2, 4)	+	−	−	+	outside
(1, 3)	+	+	−	+	outside
(0, 2)	+	+	−	−	outside
(2, 1)	0	+	+	0	vertex
(3, 2)	0	0	+	+	vertex
(2, 3)	+	0	0	+	vertex
(1, 2)	+	+	0	0	vertex

The reader may wish to verify other points, but the contents of Table 3.1 should be sufficient to demonstrate the integrity of this strategy. However, this technique is sensitive to the boundary's vertex sequence. For if the vertices are taken in a clockwise order, the four line equations are effectively multiplied by -1, which flips the normal vectors:

$$\begin{aligned} -x+y+1 &\Rightarrow x-y-1 \\ -x-y+5 &\Rightarrow x+y-5 \\ x-y+1 &\Rightarrow -x+y-1 \\ x+y-3 &\Rightarrow -x-y+3 \end{aligned}$$

Now the normal vectors point away from the boundary's inside, which makes the inside partition negative and the outside positive. This still works, but it highlights the care required when using this technique.

3.6 A line perpendicular to a vector

In this section we examine how to define a straight line that is perpendicular to a given vector and passes through a specific point. Figure 3.13 shows this scenario where the reference unit vector $\hat{\mathbf{n}}$ passes through the origin O and the specified point is $T(x_T, y_T)$.

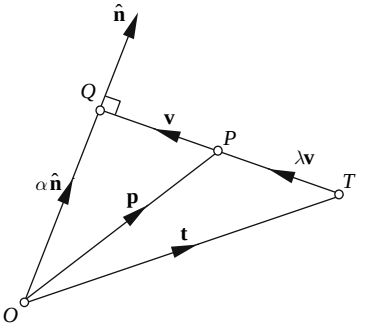


FIGURE 3.13.

First we let

$$\overrightarrow{TQ} = \mathbf{v}$$

$$\overrightarrow{TP} = \lambda\mathbf{v}$$

and

$$\overrightarrow{OQ} = \alpha\hat{\mathbf{n}}.$$

From Fig. 3.13 we can state

$$\overrightarrow{OQ} = \overrightarrow{OT} + \overrightarrow{TQ}. \tag{3.18}$$

Substituting vector names in Eq. (3.18), we obtain

$$\alpha\hat{\mathbf{n}} = \mathbf{t} + \mathbf{v}. \tag{3.19}$$

As $\hat{\mathbf{n}}$ is perpendicular to $\mathbf{v}$, we have

$$\hat{\mathbf{n}} \cdot \mathbf{v} = 0.$$

Multiplying Eq. (3.19) throughout by $\hat{\mathbf{n}}$, we obtain

$$\alpha\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = \hat{\mathbf{n}} \cdot \mathbf{t} + \hat{\mathbf{n}} \cdot \mathbf{v}.$$

Therefore,

$$\alpha\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = \hat{\mathbf{n}} \cdot \mathbf{t}$$

and

$$\alpha = \hat{\mathbf{n}} \cdot \mathbf{t}. \tag{3.20}$$

We can regard $\mathbf{p}$ as being a linear interpolation of $\mathbf{t}$ and $\alpha\hat{\mathbf{n}}$ (see Section 2.11):

$$\mathbf{p} = (1 - \lambda)\,\mathbf{t} + \lambda\alpha\hat{\mathbf{n}}. \tag{3.21}$$

Substituting Eq. (3.20) in Eq. (3.21) gives

$$\mathbf{p} = (1-\lambda)\,\mathbf{t} + \lambda\,(\hat{\mathbf{n}}\cdot\mathbf{t})\,\hat{\mathbf{n}}. \tag{3.22}$$

which is the vector equation of the line. However, there is one drawback with this equation. Notice that the point T must not lie on the perpendicular from the origin; if it does, then vector $\mathbf{v}$ would be a null vector and everything collapses!

To illustrate Eq. (3.22), let $\hat{\mathbf{n}} = \frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})$ and $T = (3, 0)$, as shown in Fig. 3.14.

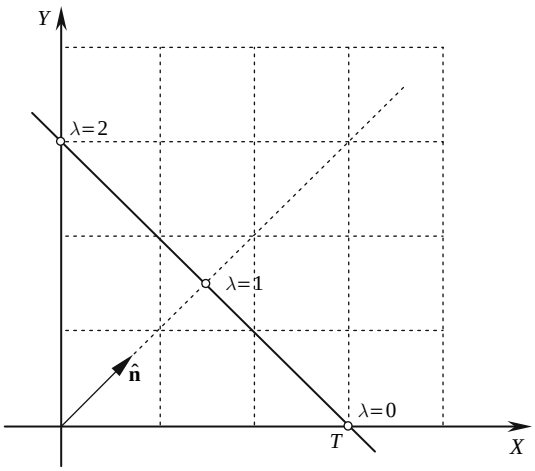


FIGURE 3.14.

When $\lambda = 0$,

$$\mathbf{p} = 3\mathbf{i} \text{ and } P = (3, 0)\,.$$

When $\lambda = 1$,

$$\mathbf{p} = \left[\tfrac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})\cdot(3\mathbf{i})\right]\tfrac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j}) = \tfrac{3}{2}(\mathbf{i}+\mathbf{j})$$

and

$$P = (1.5, 1.5)\,.$$

When $\lambda = 2$,

$$\mathbf{p} = -3\mathbf{i} + 2\left[\tfrac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})\cdot 3\mathbf{i}\right]\tfrac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j}) = -3\mathbf{i} + 3(\mathbf{i}+\mathbf{j}) = 3\mathbf{j}$$

and

$$P = (0, 3)\,.$$

This is just one way of solving this problem. There is another way, which is much simpler and exploits the perp operator.

We discovered in Section 2.10 that given $\mathbf{n} = a\mathbf{i} + b\mathbf{j}$, a perpendicular vector $\mathbf{n}^{\perp}$ exists such that $\mathbf{n}^{\perp} = -b\mathbf{i} + a\mathbf{j}$. Therefore, given a unit vector $\hat{\mathbf{n}} = x_n\mathbf{i} + y_n\mathbf{j}$, a perpendicular unit vector $\hat{\mathbf{v}}$ is given by

$$\hat{\mathbf{v}} = -y_n\mathbf{i} + x_n\mathbf{j}$$

and the line equation is

$$\mathbf{p} = \mathbf{t} + \lambda\hat{\mathbf{v}} \tag{3.23}$$

But Eq. (3.23) does not refer explicitly to the original normal vector $\hat{\mathbf{n}}$, and if this is preferred, we must find a way of transforming $x_n\mathbf{i} + y_n\mathbf{j}$ into $-y_n\mathbf{i} + x_n\mathbf{j}$. Fortunately, we can use a determinant to do this:

$$-\begin{vmatrix} \mathbf{i} & \mathbf{j} \\ x_n & y_n \end{vmatrix} = -y_n\mathbf{i} + x_n\mathbf{j}$$

Thus, the line equation is

$$\mathbf{p} = \mathbf{t} - \lambda \begin{vmatrix} \mathbf{i} & \mathbf{j} \\ x_n & y_n \end{vmatrix}. \tag{3.24}$$

Whether we use Eq. (3.23) or Eq. (3.24), we can write the line equation as

$$\begin{aligned} x_p &= x_t - \lambda y_n \\ y_p &= y_t + \lambda x_n \end{aligned} \tag{3.25}$$

Taking the previous example, where $\hat{\mathbf{n}} = \frac{1}{\sqrt{2}}(\mathbf{i} + \mathbf{j})$ and $T = (3, 0)$, we can substitute these in Eq. (3.25) and obtain

$$\begin{aligned} x_p &= 3 - \lambda\tfrac{1}{\sqrt{2}} \\ y_p &= 0 + \lambda\tfrac{1}{\sqrt{2}} \end{aligned}$$

When $\lambda = 0$,

$$P = (3, 0)\,.$$

When $\lambda = \sqrt{2}$,

$$P = (2, 1)\,.$$

3.7 The position and distance of a point on a line perpendicular to the origin

In this section we apply vector analysis to a simple geometric problem involving a straight line. Imagine, then, a straight line that does not intersect the origin. We are going to calculate the shortest distance from the origin to the line and the coordinates of the point on the line.

We examine two strategies: one for the Cartesian form of the line equation, and the other using the parametric form.

3.7.1 The Cartesian form of the line equation

With reference to Fig. 3.15, our objective is to establish an equation defining **q** in terms of the line equation.

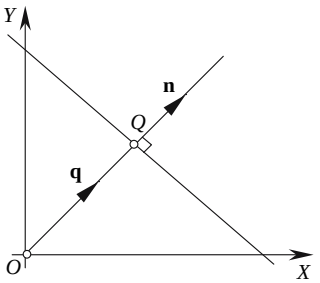


FIGURE 3.15.

We begin the process by defining the line equation as $ax + by = c$ and making Q the nearest point on the line to the origin with position vector **q**, which is perpendicular to the line. From the line equation, the vector normal to the line is

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} \tag{3.26}$$

and for the moment we will assume that **n** is not a unit vector. Let

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j}$$

Therefore,

$$\mathbf{n} \cdot \mathbf{q} = ax + by = c. \tag{3.27}$$

We now need to find the coordinates of Q and its distance from the origin.

As **q** is parallel with **n**, let

$$\mathbf{q} = \lambda \mathbf{n} \tag{3.28}$$

where λ is a scalar that has to be found to solve the problem.

Using **n**, take the dot product of Eq. (3.28) and substitute it in Eq. (3.27):

$$\mathbf{n} \cdot \mathbf{q} = \lambda \mathbf{n} \cdot \mathbf{n}$$

Therefore,

$$\lambda \mathbf{n} \cdot \mathbf{n} = c$$

where

$$\lambda = \frac{c}{\mathbf{n} \cdot \mathbf{n}} = \frac{c}{\| \mathbf{n} \|^2}. \tag{3.29}$$

By substituting λ in Eq. (3.28), we can find Q:

$$\mathbf{q} = \left(\frac{c}{\| \mathbf{n} \|^2} \right) \mathbf{n}. \tag{3.30}$$

The distance $OQ = \| \mathbf{q} \|$.

If the normal vector $\mathbf{n}$ in the line equation is a unit vector $\hat{\mathbf{n}}$, the results are greatly simplified:

$$\lambda = c$$

and

$$\mathbf{q} = c\hat{\mathbf{n}}.$$

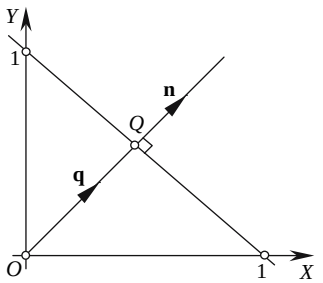


FIGURE 3.16.

Let's test Eq. (3.30) with an example.

Figure 3.16 shows the graph of straight-line equation where

$$x + y = 1$$

Therefore,

$$\mathbf{n} = \mathbf{i} + \mathbf{j} \text{ and } c = 1.$$

Using Eq. (3.29) gives

$$\lambda = \frac{c}{\| \mathbf{n} \|^2} = \tfrac{1}{2}.$$

Using Eq. (3.30) shows that

$$\mathbf{q} = \tfrac{1}{2} (\mathbf{i} + \mathbf{j}).$$

The nearest point to O is $Q(1/2, 1/2)$, and the distance

$$OQ = \| \mathbf{q} \| = \tfrac{1}{2} \| \mathbf{n} \| = \tfrac{1}{2}\sqrt{2} = 0.7071.$$

Now let's examine the same problem using the parametric form of the line equation.

3.7.2 The parametric form of the line equation

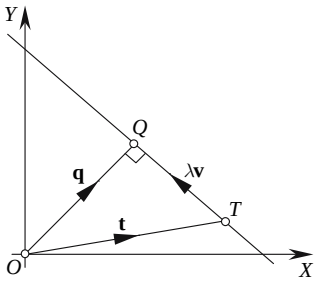


FIGURE 3.17.

With reference to Fig. 3.17, let the straight-line equation be

$$\mathbf{q} = \mathbf{t} + \lambda\mathbf{v}. \tag{3.31}$$

The objective is to establish the value of λ.

Q is nearest to O when $\mathbf{q}$ is perpendicular to. Therefore,

$$\mathbf{v} \cdot \mathbf{q} = 0. \tag{3.32}$$

Using $\mathbf{v}$, take the dot product of Eq. (3.31) and substitute it in Eq. (3.32):

$$\mathbf{v} \cdot \mathbf{q} = \mathbf{v} \cdot \mathbf{t} + \lambda\mathbf{v} \cdot \mathbf{v} = 0$$

Therefore,

$$\lambda = \frac{-\mathbf{v} \cdot \mathbf{t}}{\mathbf{v} \cdot \mathbf{v}} = \frac{-\mathbf{v} \cdot \mathbf{t}}{\| \mathbf{v} \|^2}.$$

The position vector for Q is then given by

$$\mathbf{q} = \mathbf{t} - \left(\frac{\mathbf{v} \cdot \mathbf{t}}{\| \mathbf{v} \|^2} \right) \mathbf{v} \tag{3.33}$$

and the distance $OQ = \| \mathbf{q} \|$.

Eq. (3.33) is greatly simplified if $\mathbf{v}$ is a unit vector $\hat{\mathbf{v}}$:

$$\mathbf{q} = \mathbf{t} + \lambda\hat{\mathbf{v}}.$$

Therefore,

$$\lambda = -\hat{\mathbf{v}} \cdot \mathbf{t}$$

and

$$\mathbf{q} = \mathbf{t} - (\hat{\mathbf{v}} \cdot \mathbf{t})\,\hat{\mathbf{v}}. \tag{3.34}$$

Again, let's test Eq. (3.34) with an example.

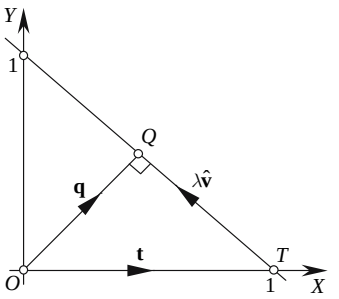


FIGURE 3.18.

With reference to Fig. 3.18, let the line equation be

$$\mathbf{q} = \mathbf{t} + \lambda \hat{\mathbf{v}}$$

where

$$\mathbf{t} = \mathbf{i} \text{ and } \hat{\mathbf{v}} = \tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j}).$$

Therefore,

$$\lambda = -\tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j}) \cdot \mathbf{i} = \tfrac{1}{\sqrt{2}}$$

and

$$\mathbf{q} = \mathbf{i} + \tfrac{1}{\sqrt{2}}\left(\tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j})\right) = \mathbf{i} - \tfrac{1}{2}\mathbf{i} + \tfrac{1}{2}\mathbf{j}$$
$$\mathbf{q} = \tfrac{1}{2}\mathbf{i} + \tfrac{1}{2}\mathbf{j}.$$

Therefore,

$$Q = \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \text{ and } OQ = \sqrt{\tfrac{1}{2}^2 + \tfrac{1}{2}^2} = \tfrac{1}{\sqrt{2}} = 0.7071.$$

3.8 The position and distance of the nearest point on a line to a point

In computer graphics we are often interested in knowing the distance between two objects, which depends on whether the objects are represented by polygons, patches, or functions. To begin this analysis, let's start with finding the distance between the nearest point on a line to a point.

As we saw in Section 3.7, the form of the line equation gives rise to two solutions, so we will explore both the Cartesian and parametric forms.

3.8.1 The Cartesian form of the line equation

The strategy is virtually identical to that used in Section 3.7, apart from the fact that the point is not located at the origin.

Figure 3.19 shows a typical scenario with the point P and its closest neighbor $Q(x, y)$ on the line. There are four vectors to consider:

$\mathbf{n}$ is the line's normal vector,
$\mathbf{p}$ is P's position vector,
$\mathbf{q}$ is Q's position vector, and
$\lambda\mathbf{n}$ connects P to Q, where λ is a scalar.

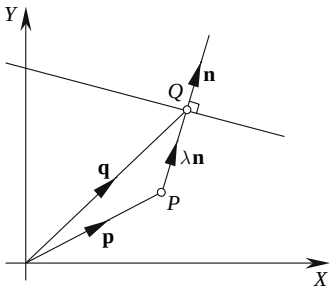


FIGURE 3.19.

With reference to Fig. 3.19, let the line equation be

$$ax + by = c.$$

Then the vector normal to the line is

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j}.$$

Let

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j}.$$

Therefore,

$$\mathbf{n} \cdot \mathbf{q} = ax + by = c. \tag{3.35}$$

From Fig. 3.19, we have

$$\mathbf{q} = \mathbf{p} + \lambda\mathbf{n}. \tag{3.36}$$

The objective is to find the value of λ.
Multiply Eq. (3.36) throughout using $\mathbf{n}$:

$$\mathbf{n} \cdot \mathbf{q} = \mathbf{n} \cdot \mathbf{p} + \lambda\mathbf{n} \cdot \mathbf{n}. \tag{3.37}$$

Substituting Eq. (3.35) in Eq. (3.37), we get

$$c = \mathbf{n} \cdot \mathbf{p} + \lambda \mathbf{n} \cdot \mathbf{n}$$

which enables us to define λ:

$$\lambda = \frac{c - \mathbf{n} \cdot \mathbf{p}}{\mathbf{n} \cdot \mathbf{n}} = \frac{c - \mathbf{n} \cdot \mathbf{p}}{\| \mathbf{n} \|^2}. \tag{3.38}$$

Substituting λ in Eq. (3.36) gives

$$\mathbf{q} = \mathbf{p} + \left(\frac{c - \mathbf{n} \cdot \mathbf{p}}{\| \mathbf{n} \|^2} \right) \mathbf{n}. \tag{3.39}$$

The distance

$$PQ = \| \lambda \mathbf{n} \| .$$

If $\mathbf{n}$ is a unit vector $\hat{\mathbf{n}}$, Eq. (3.39) simplifies to

$$\mathbf{q} = \mathbf{p} + (c - \hat{\mathbf{n}} \cdot \mathbf{p}) \hat{\mathbf{n}}$$

and the distance

$$PQ = \| c - \hat{\mathbf{n}} \cdot \mathbf{p} \| .$$

Equation (3.39) covers all positions for P, including the origin. Therefore, if we make $\mathbf{p}$ a null vector,

$$\mathbf{q} = \left(\frac{c}{\| \mathbf{n} \|^2} \right) \mathbf{n}$$

which is identical to Eq. (3.30)!

Let's test Eq. (3.39) with a predictable example, as shown in Fig. 3.20.

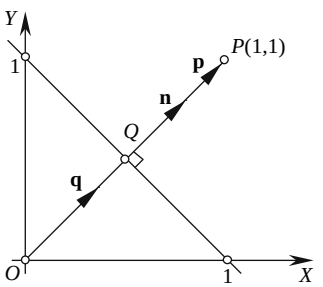


Figure 3.20.

P has coordinates (1, 1), which means that Q has coordinates $\left(\frac{1}{2}, \frac{1}{2}\right)$:

$$\mathbf{p} = \mathbf{i} + \mathbf{j}.$$

The line equation is

$$x + y = 1.$$

Therefore,

$$\mathbf{n} = \mathbf{i} + \mathbf{j}$$

and

$$c = 1.$$

Using Eq. (3.38),

$$\lambda = \frac{c - \mathbf{n} \cdot \mathbf{p}}{\mathbf{n} \cdot \mathbf{n}} = \frac{1-2}{2} = -\tfrac{1}{2}.$$

Using Eq. (3.39),

$$\mathbf{q} = \mathbf{p} - \tfrac{1}{2}\mathbf{n} = (\mathbf{i}+\mathbf{j}) - \tfrac{1}{2}(\mathbf{i}+\mathbf{j}) = \tfrac{1}{2}\mathbf{i} + \tfrac{1}{2}\mathbf{j}$$

giving

$$Q\left(\tfrac{1}{2},\ \tfrac{1}{2}\right).$$

The distance

$$PQ = \| \lambda \mathbf{n} \| = \left\| -\tfrac{1}{2}(\mathbf{i}+\mathbf{j}) \right\| = \tfrac{\sqrt{2}}{2} = 0.7071.$$

Now let's consider the parametric form of the line equation.

3.8.2 The parametric form of the line equation

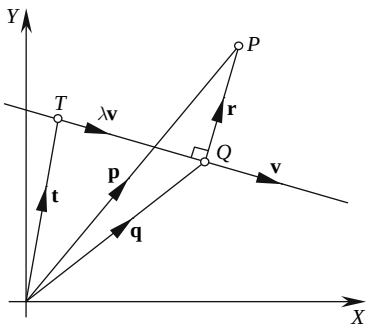


FIGURE 3.21.

With reference to Fig. 3.21, let the straight-line equation be

$$\mathbf{q} = \mathbf{t} + \lambda \mathbf{v}. \tag{3.40}$$

The objective is to find the value of λ.

Q is the nearest point on the line to P with position vector $\mathbf{q}$.

Because $\mathbf{r} \perp \mathbf{v}$, $\mathbf{p}$, and $\mathbf{q}$ have identical projections on $\mathbf{v}$, we find that

$$\mathbf{v} \cdot \mathbf{p} = \mathbf{v} \cdot \mathbf{q}. \tag{3.41}$$

Multiply Eq. (3.40) throughout using **v**:

$$\mathbf{v}\cdot\mathbf{q} = \mathbf{v}\cdot\mathbf{t} + \lambda\mathbf{v}\cdot\mathbf{v}. \tag{3.42}$$

Substituting Eq. (3.41) in Eq. (3.42) gives

$$\mathbf{v}\cdot\mathbf{p} = \mathbf{v}\cdot\mathbf{t} + \lambda\mathbf{v}\cdot\mathbf{v}. \tag{3.43}$$

From Eq. (3.43), we see that

$$\lambda = \frac{\mathbf{v}\cdot(\mathbf{p}-\mathbf{t})}{\mathbf{v}\cdot\mathbf{v}} = \frac{\mathbf{v}\cdot(\mathbf{p}-\mathbf{t})}{\|\mathbf{v}\|^2}. \tag{3.44}$$

Substituting Eq. (3.44) in Eq. (3.40) gives

$$\mathbf{q} = \mathbf{t} + \left(\frac{\mathbf{v}\cdot(\mathbf{p}-\mathbf{t})}{\|\mathbf{v}\|^2}\right)\mathbf{v}. \tag{3.45}$$

The distance $PQ = \|\mathbf{r}\| = \|\mathbf{p}-\mathbf{q}\| = \|\mathbf{p}-(\mathbf{t}+\lambda\mathbf{v})\|$.

If **v** is a unit vector $\hat{\mathbf{v}}$, then

$$\lambda = \hat{\mathbf{v}}\cdot(\mathbf{p}-\mathbf{t})$$

$$\mathbf{q} = \mathbf{t} + (\hat{\mathbf{v}}\cdot(\mathbf{p}-\mathbf{t}))\,\hat{\mathbf{v}}.$$

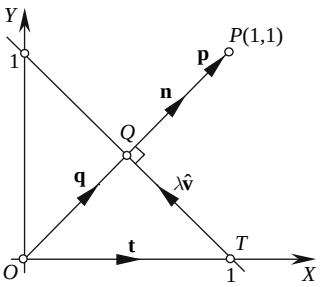


Figure 3.22.

Let's test the validity of Eq. (3.45).

Figure 3.22 shows a parametric line and a point $P(1, 1)$. The line equation is

$$\mathbf{q} = \mathbf{t} + \lambda\hat{\mathbf{v}}$$

where

$$\hat{\mathbf{v}} = \tfrac{1}{\sqrt{2}}(-\mathbf{i}+\mathbf{j}).$$

Also,

$$\mathbf{p} = \mathbf{i}+\mathbf{j}$$

and

$$\mathbf{t} = \mathbf{i}.$$

Therefore,

$$\lambda = \hat{\mathbf{v}} \cdot (\mathbf{p} - \mathbf{t}) = \tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j}) \cdot (\mathbf{i} + \mathbf{j} - \mathbf{i}) = \tfrac{1}{\sqrt{2}}$$

and

$$\mathbf{q} = \mathbf{i} + \tfrac{1}{\sqrt{2}}\tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j}) = \tfrac{1}{2}\mathbf{i} + \tfrac{1}{2}\mathbf{j}.$$

The coordinates of Q are $\left(\tfrac{1}{2}, \tfrac{1}{2}\right)$.

The distance

$$PQ = \left\| \mathbf{i} + \mathbf{j} - \left(\mathbf{i} + \tfrac{1}{\sqrt{2}}\tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j})\right) \right\| = \left\| \tfrac{1}{2}\mathbf{i} + \tfrac{1}{2}\mathbf{j} \right\| = \tfrac{\sqrt{2}}{2} = 0.7071.$$

3.9 The position of a point reflected in a line

In this section we consider the problem of calculating the reflection of a point in a straight line, and once more we consider two separate strategies for two forms of the line equation. In both cases, the key to the solution is based on the fact that the line connecting the point to its reflection is perpendicular to the reflecting line.

3.9.1 The Cartesian form of the line equation

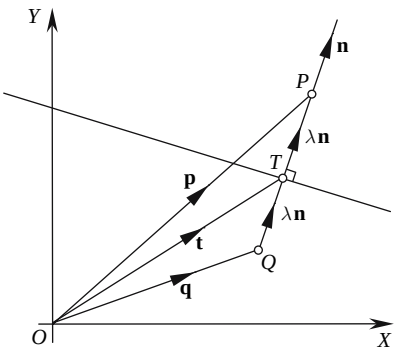


Figure 3.23.

The objective is to produce a vector equation of the form $\mathbf{q} = \mathbf{p} + \ldots$. The RHS of this equation can only contain references to $\mathbf{p}$ and variables in the line equation. λ is acceptable, so long as it can be defined. As $\overrightarrow{QP}$ is orthogonal to the reflecting line, we must anticipate employing the dot product.

With reference to Fig. 3.23, let the line equation be

$$ax + by = c.$$

Then the vector normal to the line is

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j}.$$

Let $T(x, y)$ be a point on the line with position vector $\mathbf{t} = x\mathbf{i} + y\mathbf{j}$. Therefore,

$$\mathbf{n} \cdot \mathbf{t} = ax + by = c. \tag{3.46}$$

Now

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{n}. \tag{3.47}$$

Using $\mathbf{n}$, take the dot product of Eq. (3.47):

$$\mathbf{n} \cdot \mathbf{p} = \mathbf{n} \cdot \mathbf{t} + \lambda\mathbf{n} \cdot \mathbf{n}. \tag{3.48}$$

Substituting Eq. (3.46) in Eq. (3.48) gives

$$\mathbf{n} \cdot \mathbf{p} = c + \lambda\mathbf{n} \cdot \mathbf{n}.$$

Therefore, we have

$$\lambda = \frac{\mathbf{n} \cdot \mathbf{p} - c}{\mathbf{n} \cdot \mathbf{n}} = \frac{\mathbf{n} \cdot \mathbf{p} - c}{\| \mathbf{n} \|^2}. \tag{3.49}$$

Now

$$\mathbf{q} = \mathbf{p} - 2\lambda\mathbf{n}.$$

Therefore,

$$\mathbf{q} = \mathbf{p} - 2\left(\frac{\mathbf{n} \cdot \mathbf{p} - c}{\| \mathbf{n} \|^2}\right)\mathbf{n}. \tag{3.50}$$

If $\mathbf{n}$ is a unit vector $\hat{\mathbf{n}}$, then

$$\mathbf{q} = \mathbf{p} - 2\left(\hat{\mathbf{n}} \cdot \mathbf{p} - c\right)\hat{\mathbf{n}}. \tag{3.51}$$

Before proceeding with the parametric form of the line equation, let's test Eq. (3.51) with a simple example — one where we can predict the result.

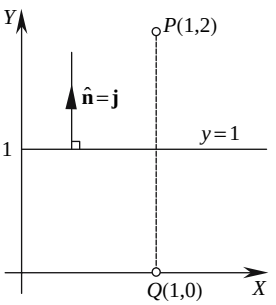


Figure 3.24.

Figure 3.24 shows the graph of the line equation $y = 1$. Thus, $\hat{\mathbf{n}} = \mathbf{j}$ and $c = 1$. The reference point is $P(1, 2)$, and it is obvious that the reflected point must be $Q(1, 0)$.

Using Eq. (3.50), we obtain

$$\mathbf{q} = \mathbf{p} - 2(\hat{\mathbf{n}} \cdot \mathbf{p} - c)\hat{\mathbf{n}} = (\mathbf{i} + \mathbf{j}) - 2(\mathbf{j} \cdot (\mathbf{i} + \mathbf{j}) - 1)\mathbf{j}$$

$$\mathbf{q} = (\mathbf{i} + 2\mathbf{j}) - 2(\mathbf{j} \cdot (\mathbf{i} + 2\mathbf{j}) - 1)\mathbf{j} = \mathbf{i}.$$

Therefore,

$$Q = (1, 0).$$

3.9.2 The parametric form of the line equation

Figure 3.25 illustrates a line whose direction is given by $\mathbf{v}$. $T(x, y)$is some point on the line with position vector $\mathbf{t} = x\mathbf{i} + y\mathbf{j}$, and P is some arbitrary point with position vector $\mathbf{p}$. Q is the reflection of P with position vector $\mathbf{q}$.

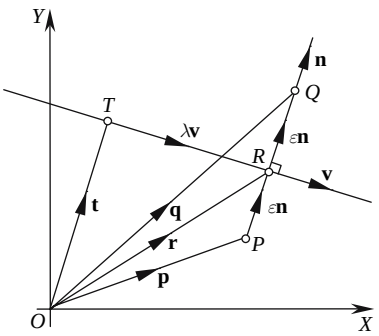


Figure 3.25.

The following analysis exploits the fact that $\overrightarrow{PQ}$ is perpendicular to $\mathbf{v}$, and the objective is to find $\mathbf{q}$ in terms of $\mathbf{t}$, $\mathbf{p}$, and $\mathbf{v}$. We begin with

$$\mathbf{r} = \mathbf{t} + \lambda\mathbf{v}. \tag{3.52}$$

Using $\mathbf{v}$, take the dot product of Eq. (3.52):

$$\mathbf{v} \cdot \mathbf{r} = \mathbf{v} \cdot \mathbf{t} + \lambda\mathbf{v} \cdot \mathbf{v}. \tag{3.53}$$

As $\mathbf{p}$ and $\mathbf{r}$ have a common projection on $\mathbf{v}$, we have

$$\mathbf{v} \cdot \mathbf{p} = \mathbf{v} \cdot \mathbf{r}. \tag{3.54}$$

Substituting Eq. (3.54) in Eq. (3.53) gives

$$\mathbf{v} \cdot \mathbf{p} = \mathbf{v} \cdot \mathbf{t} + \lambda\mathbf{v} \cdot \mathbf{v}.$$

Therefore,

$$\lambda = \frac{\mathbf{v} \cdot (\mathbf{p} - \mathbf{t})}{\mathbf{v} \cdot \mathbf{v}} = \frac{\mathbf{v} \cdot (\mathbf{p} - \mathbf{t})}{\|\mathbf{v}\|^2}. \tag{3.55}$$

As $\mathbf{r}$ is mid-way between $\mathbf{p}$ and $\mathbf{q}$, we have

$$\mathbf{r} = \tfrac{1}{2}\mathbf{p} + \tfrac{1}{2}\mathbf{q}.$$

and

$$\mathbf{q} = 2\mathbf{r} - \mathbf{p}. \tag{3.56}$$

Substituting Eq. (3.52) in Eq. (3.56) gives

$$\mathbf{q} = 2(\mathbf{t} + \lambda\mathbf{v}) - \mathbf{p}.$$

If $\mathbf{v}$ is a unit vector $\hat{\mathbf{v}}$, then

$$\lambda = \hat{\mathbf{v}} \cdot (\mathbf{p} - \mathbf{t}) \tag{3.57}$$

$$\mathbf{q} = 2(\mathbf{t} + \lambda\hat{\mathbf{v}}) - \mathbf{p} \tag{3.58}$$

Let's test Eq. (3.4) with a simple example.

Figure 3.26 illustrates a line whose direction is given by $\hat{\mathbf{v}} = \frac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j})$. $T(1, 0)$ is some point on the line with position vector $\mathbf{t} = \mathbf{i}$, and $P(1, 1)$ is some arbitrary point with position vector $\mathbf{p} = \mathbf{i} + \mathbf{j}$.

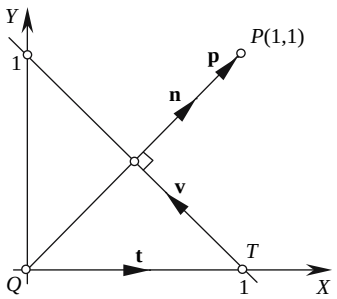


Figure 3.26.

Using Eq. (3.57) gives

$$\lambda = \hat{\mathbf{v}} \cdot (\mathbf{p} - \mathbf{t}) = \tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j}) \cdot (\mathbf{i} + \mathbf{j} - \mathbf{i}) = \tfrac{1}{\sqrt{2}}$$

$$\lambda = \tfrac{1}{\sqrt{2}}.$$

Using Eq. (3.58) gives

$$\mathbf{q} = 2(\mathbf{t} + \lambda\hat{\mathbf{v}}) - \mathbf{p} = 2\left(\mathbf{i} + \tfrac{1}{\sqrt{2}}\tfrac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j})\right) - \mathbf{i} - \mathbf{j} = 2\mathbf{i} - \mathbf{i} + \mathbf{j} - \mathbf{i} - \mathbf{j} = 0.$$

Therefore, the reflected point is $Q(0, 0)$.

3.10 A line perpendicular to a line through a point

Here is another geometric problem. Given a line m and a point P, what is the equation of a straight line that passes through P and is perpendicular to m? The algebraic solution is relatively easy to solve, so let's consider a vector solution.

3.10.1 The parametric form of the line equation

For 2D lines we can exploit the perp operator.

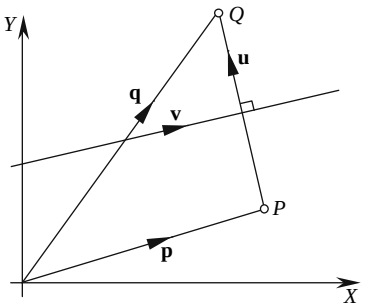


FIGURE 3.27.

Let the line's direction vector be $\mathbf{v}$; therefore, the perpendicular vector is $\mathbf{u} = \mathbf{v}^{\perp}$ (Fig. 3.27).

Therefore, the equation of the perpendicular line is

$$\mathbf{q} = \mathbf{p} + \lambda \mathbf{u}$$

or

$$\mathbf{q} = \mathbf{p} + \lambda \mathbf{v}^{\perp}.$$

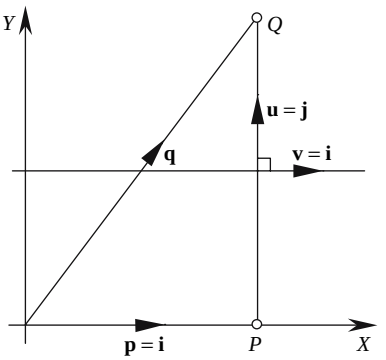


FIGURE 3.28.

For example, Fig. 3.28 shows a 2D line where

$$\mathbf{v} = \mathbf{i} \text{ and } \mathbf{u} = \mathbf{j}.$$

If the point P is $(1, 0)$, then $\mathbf{p} = \mathbf{i}$ and

$$\mathbf{q} = \mathbf{p} + \lambda \mathbf{v}^{\perp}$$

$$\mathbf{q} = \mathbf{i} + \lambda \mathbf{j}.$$

Now let's consider the solution for a 2D or 3D line.

We begin with the scenario shown in Fig. 3.29.

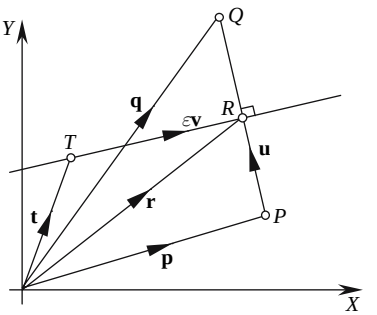


FIGURE 3.29.

The objective is to find an equation defining $\mathbf{q}$ in terms of $\mathbf{p}$, $\mathbf{v}$, and $\mathbf{t}$. We start with

$$\mathbf{u} = \mathbf{r} - \mathbf{p}.$$

Let

$$\mathbf{r} = \mathbf{t} + \varepsilon\mathbf{v}, \text{ where } \varepsilon \text{ is a scalar.} \tag{3.59}$$

Therefore,

$$\mathbf{u} = \mathbf{t} + \varepsilon\mathbf{v} - \mathbf{p}.$$

Because $\mathbf{u}$ is perpendicular to $\mathbf{v}$, $\mathbf{r}$ and $\mathbf{p}$ have the same projection on $\mathbf{v}$. Therefore,

$$\mathbf{v} \cdot \mathbf{r} = \mathbf{v} \cdot \mathbf{p}. \tag{3.60}$$

Using $\mathbf{v}$, take the dot product of Eq. (3.59):

$$\mathbf{v} \cdot \mathbf{r} = \mathbf{v} \cdot \mathbf{t} + \varepsilon\mathbf{v} \cdot \mathbf{v}. \tag{3.61}$$

Substituting Eq. (3.60) in Eq. (3.61) gives

$$\mathbf{v} \cdot \mathbf{p} = \mathbf{v} \cdot \mathbf{t} + \varepsilon\mathbf{v} \cdot \mathbf{v}.$$

Therefore,

$$\varepsilon = \frac{\mathbf{v} \cdot (\mathbf{p} - \mathbf{t})}{\mathbf{v} \cdot \mathbf{v}} = \frac{\mathbf{v} \cdot (\mathbf{p} - \mathbf{t})}{\| \mathbf{v} \|^2}. \tag{3.62}$$

Therefore, we can write

$$\mathbf{q} = \mathbf{p} + \lambda\mathbf{u}$$

where

$$\mathbf{u} = \mathbf{t} + \varepsilon\mathbf{v} - \mathbf{p}.$$

If $\mathbf{v}$ is a unit vector $\hat{\mathbf{v}}$, then

$$\varepsilon = \hat{\mathbf{v}} \cdot (\mathbf{p} - \mathbf{t}) \tag{3.63}$$

$$\mathbf{u} = \mathbf{t} + \varepsilon\hat{\mathbf{v}} - \mathbf{p} \tag{3.64}$$

and

$$\mathbf{q} = \mathbf{p} + \lambda\mathbf{u}. \tag{3.65}$$

Let's find $\mathbf{u}$ in Fig. 3.30.

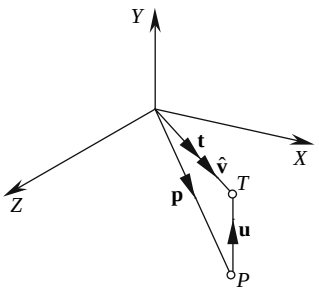


FIGURE 3.30.

It just so happens that the values of $\mathbf{p}$ and $\mathbf{t}$ give rise to a simple solution.
Nevertheless, we will apply Eqs. (3.63) and (3.64) to demonstrate their validity.
Let

$$\hat{\mathbf{v}} = \tfrac{1}{\sqrt{2}}(\mathbf{i} + \mathbf{k})$$

$$\mathbf{t} = \mathbf{i} + \mathbf{k}$$

$$\mathbf{p} = \mathbf{i} - \mathbf{j} + \mathbf{k}.$$

Then

$$\varepsilon = \tfrac{1}{\sqrt{2}}(\mathbf{i} + \mathbf{k}) \cdot (\mathbf{i} - \mathbf{j} + \mathbf{k} - \mathbf{i} - \mathbf{k}) = \tfrac{1}{\sqrt{2}}(\mathbf{i} + \mathbf{k}) \cdot (-\mathbf{j}) = 0$$

$$\varepsilon = 0.$$

Using Eq. (3.64) gives

$$\mathbf{u} = \mathbf{i} + \mathbf{k} - \mathbf{i} + \mathbf{j} - \mathbf{k} = \mathbf{j}.$$

Therefore, the line equation is $\mathbf{q} = \mathbf{p} + \lambda\mathbf{j}$, which is correct.

3.11 A line equidistant from two points

The problem here is to identify a line that is equidistant from two points and is perpendicular to the line connecting the two points. This is a problem that only exists in R^2; the R^3 equivalent involves a plane equidistant from two points. So let's consider a parametric solution using the perp operator.

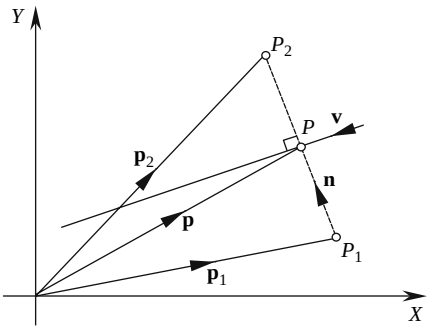


Figure 3.31.

With reference to Fig. 3.31, we see that

$$\mathbf{n} = \mathbf{p}_2 - \mathbf{p}_1 = (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\,\mathbf{j}.$$

The line's direction vector $\mathbf{v}$ is perpendicular to $\mathbf{n}$:

$$\mathbf{v} = \mathbf{n}^{\perp} = -(y_2 - y_1)\,\mathbf{i} + (x_2 - x_1)\,\mathbf{j}.$$

The position vector $\mathbf{p}$ is given by

$$\mathbf{p} = \tfrac{1}{2}(\mathbf{p}_2 + \mathbf{p}_1).$$

Therefore, any point $\mathbf{q}$ on the line is given by

$$\mathbf{q} = \mathbf{p} + \lambda\mathbf{v}.$$

More explicitly:

$$x_q = \tfrac{1}{2}(x_2 + x_1) - \lambda(y_2 - y_1)$$
$$y_q = \tfrac{1}{2}(y_2 + y_1) + \lambda(x_2 - x_1).$$

3.12 The equation of a line segment

The line segment is very important in computer graphics as it is a fundamental building block for 2D and 3D polygons, and all sorts of surface meshes. In R^2 we need to clip line segments against rectangular windows; in R^3 we need to clip line segments against a viewing frustum; and in R^2 and R^3 we need to compute the potential intersection of two line segments.

The parametric form of the straight-line equation is the most practical basis for manipulating straight-line segments, where the controlling parameter can be used to determine the position of a point along the segment.

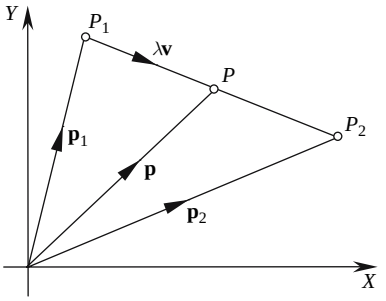


FIGURE 3.32.

With reference to Fig. 3.32, we define a line segment by two points $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ and their respective position vectors $\mathbf{p}_1$ and $\mathbf{p}_2$. $P(x_P, y_P)$ is a point on the line segment with position vector $\mathbf{p}$.

Let

$$\mathbf{v} = \mathbf{p}_2 - \mathbf{p}_1$$

and

$$\mathbf{p} = \mathbf{p}_1 + \lambda\mathbf{v}.$$

Therefore,

$$x_P = x_1 + \lambda(x_2 - x_1) \qquad y_P = y_1 + \lambda(y_2 - y_1)$$

P is between P_1 and P_2 for $\lambda \in [0, 1]$.

By changing λ between 0 and 1, we slide along the line segment.

3.13 The intersection of two straight lines

The need to compute line intersections arises in 2D and 3D clipping, modeling, and animation. Most of the time we are concerned with line segments, rather than infinite lines, although the latter are required in illumination calculations, ray casting, and ray tracing. To set the scene, let us calculate the point of intersection of two straight lines in R^2.

To begin with, we must anticipate all the possibilities that could arise with two lines: the first is that the lines do intersect, giving a single point of intersection; second, the lines could be coincident, producing an infinite number of intersections; and third, the lines are distinct, but parallel, producing no intersections. Although one can approach this problem using line equations, we consider only a vector approach.

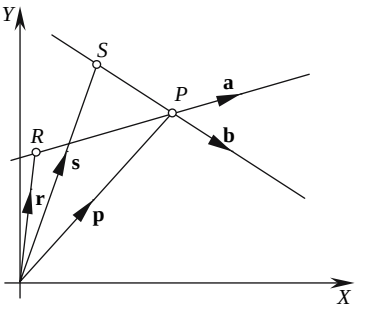


Figure 3.33.

The lines are shown in Fig. 3.33 and are defined as follows:

$$\mathbf{p} = \mathbf{r} + \lambda\mathbf{a}$$
$$\mathbf{p} = \mathbf{s} + \varepsilon\mathbf{b}$$

where λ and ε are scalars, and the direction vectors are

$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j}$$
$$\mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j}.$$

The task now is to discover the values of λ and ε at the point P. Let $P(x_P, y_P)$ be the point of intersection of the two lines and $\mathbf{p}$ its position vector. Therefore,

$$\mathbf{r} + \lambda\mathbf{a} = \mathbf{s} + \varepsilon\mathbf{b} \tag{3.66}$$

In order to isolate λ, we multiply Eq. (3.71) by $\mathbf{b}^\perp$:

$$\lambda\mathbf{b}^\perp \cdot \mathbf{a} = \mathbf{b}^\perp \cdot (\mathbf{s} - \mathbf{r}) + \varepsilon\mathbf{b}^\perp \cdot \mathbf{b}$$

and as $\mathbf{b}^\perp \cdot \mathbf{b} = 0$,

$$\lambda = \frac{\mathbf{b}^\perp \cdot (\mathbf{s} - \mathbf{r})}{\mathbf{b}^\perp \cdot \mathbf{a}}$$

from which we can state that

$$\lambda = \frac{x_b(y_S - y_R) - y_b(x_S - x_R)}{x_b y_a - x_a y_b}.$$

In order to isolate ε, we multiply Eq. (3.71) by $\mathbf{a}^{\perp}$:

$$\varepsilon \mathbf{a}^{\perp} \cdot \mathbf{b} = \mathbf{a}^{\perp} \cdot (\mathbf{r} - \mathbf{s}) + \lambda \mathbf{a}^{\perp} \cdot \mathbf{a}$$

and as $\mathbf{a}^{\perp} \cdot \mathbf{a} = 0$,

$$\varepsilon \mathbf{a}^{\perp} \cdot \mathbf{b} = \mathbf{a}^{\perp} \cdot (\mathbf{r} - \mathbf{s})$$

$$\varepsilon = \frac{\mathbf{a}^{\perp} \cdot (\mathbf{r} - \mathbf{s})}{\mathbf{a}^{\perp} \cdot \mathbf{b}}$$

from which we can state that

$$\varepsilon = \frac{-y_a (x_R - x_S) + x_a (y_R - y_S)}{-y_a x_b + x_a y_b}$$

or

$$\varepsilon = \frac{x_a (y_S - y_R) - y_a (x_S - x_R)}{x_b y_a - x_a y_b}.$$

The coordinates of P are given by

$$x_P = x_R + \lambda x_a \qquad y_P = y_R + \lambda y_a$$

or

$$x_P = x_S + \varepsilon x_b \qquad y_P = y_S + \varepsilon y_b.$$

But what if the lines are parallel? Well, this can be detected when $\mathbf{a} = t\mathbf{b}$, i.e., $x_b y_a - x_a y_b = 0$.

Before moving on to the intersection of two line segments, let's test the above equations with an example.

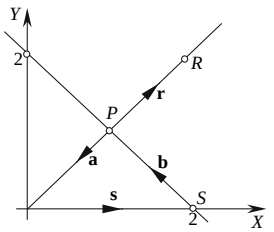


FIGURE 3.34.

Figure 3.34 shows two lines intersecting at P. The line equations are given by

$$\mathbf{r} = 2\mathbf{i} + 2\mathbf{j} \text{ and } \mathbf{a} = -\mathbf{i} - \mathbf{j}$$

$$\mathbf{s} = 2\mathbf{i} \text{ and } \mathbf{b} = -\mathbf{i} + \mathbf{j}.$$

Therefore,

$$\lambda = \frac{-1(-2)-(0)}{(-1\times -1)-(-1\times 1)} = 1$$

or

$$\varepsilon = \frac{-1(-2)+1(0)}{-1(-1)+1(1)} = 1$$

and

$$x_P = x_R + \lambda x_a = 2-1 = 1$$
$$y_P = y_R + \lambda y_a = 2-1 = 1$$

making the intersection point (1, 1), which is correct.

3.14 The point of intersection of two 2D line segments

In Section 3.13 we calculated the intersection point of two infinite 2D lines, and in this section we exploit this knowledge to detect the potential intersection of two finite 2D line segments. As the segments are represented by parametric vectors, the value of the parameter is the key to identifying intersections, non-intersections, and touch conditions.

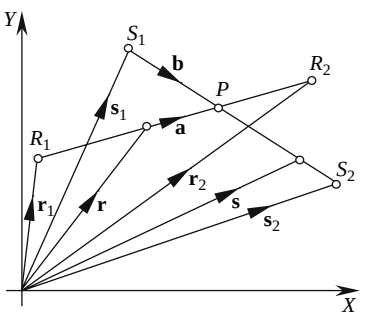


Figure 3.35.

We let the end points of the line segments be defined by $R_1(x_{R1}, y_{R1})$, $R_2(x_{R2}, y_{R2})$ and $S_1(x_{S1}, y_{S1})$, $S_2(x_{S2}, y_{S2})$, where $P(x_P, y_P)$ is the point of intersection. From Fig. 3.35, we see that

$$\mathbf{a} = \mathbf{r}_2 - \mathbf{r}_1 \text{ and } \mathbf{b} = \mathbf{s}_2 - \mathbf{s}_1$$

and the line segment equations are

$$\mathbf{r} = \mathbf{r}_1 + \lambda\mathbf{a} \text{ and } \mathbf{s} = \mathbf{s}_1 + \varepsilon\mathbf{b}.$$

For intersection at P, we have

$$\mathbf{r}_1 + \lambda\mathbf{a} = \mathbf{s}_1 + \varepsilon\mathbf{b}$$

where

$$\lambda\mathbf{a} = \mathbf{s}_1 - \mathbf{r}_1 + \varepsilon\mathbf{b} \tag{3.67}$$

and

$$\varepsilon\mathbf{b} = \mathbf{r}_1 - \mathbf{s}_1 + \lambda\mathbf{a}. \tag{3.68}$$

To find λ, multiply Eq. (3.67) by $\mathbf{b}^{\perp}$:

$$\lambda\mathbf{b}^{\perp} \cdot \mathbf{a} = \mathbf{b}^{\perp} \cdot (\mathbf{s}_1 - \mathbf{r}_1) + \varepsilon\mathbf{b}^{\perp} \cdot \mathbf{b}$$

and as $\mathbf{b}^{\perp} \cdot \mathbf{b} = 0$,

$$\lambda = \frac{\mathbf{b}^{\perp} \cdot (\mathbf{s}_1 - \mathbf{r}_1)}{\mathbf{b}^{\perp} \cdot \mathbf{a}}$$

and

$$\lambda = \frac{x_b (y_{S1} - y_{R1}) - y_b (x_{S1} - x_{R1})}{x_b y_a - y_b x_a}. \tag{3.69}$$

To find ε, multiply Eq. (3.68) by $\mathbf{a}^{\perp}$:

$$\varepsilon\mathbf{a}^{\perp} \cdot \mathbf{b} = \mathbf{a}^{\perp} \cdot (\mathbf{r}_1 - \mathbf{s}_1) + \lambda\mathbf{a}^{\perp} \cdot \mathbf{a}$$

and as $\mathbf{a}^{\perp} \cdot \mathbf{a} = 0$,

$$\varepsilon = \frac{\mathbf{a}^{\perp} \cdot (\mathbf{r}_1 - \mathbf{s}_1)}{\mathbf{a}^{\perp} \cdot \mathbf{b}}$$

and

$$\varepsilon = \frac{x_a (y_{R1} - y_{S1}) - y_a (x_{R1} - x_{S1})}{x_a y_b - y_a x_b}. \tag{3.70}$$

If $\lambda, \varepsilon \in [0, 1]$, the lines intersect or touch one another, and the coordinates of P are given by

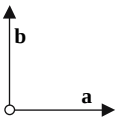
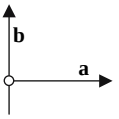
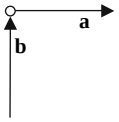
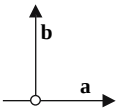
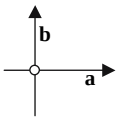
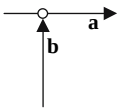
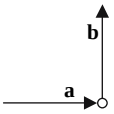
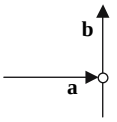
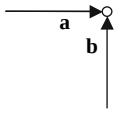
$$x_P = x_{R1} + \lambda x_a \qquad y_P = y_{R1} + \lambda y_a$$

or

$$x_P = x_{S1} + \varepsilon x_b \qquad y_P = y_{S1} + \varepsilon y_b.$$

The line segments are parallel if the denominator in Eq. (3.69) or Eq. (3.70) is zero. Table 3.2 illustrates the relative positions of the line segments for different values of λ and ε.

TABLE 3.2

λ	ε		ε		ε	
0	0		$0 < \varepsilon < 1$		1	
$0 < \lambda < 1$	0		$0 < \varepsilon < 1$		1	
1	0		$0 < \varepsilon < 1$		1	

4 The Plane

4.1 Introduction

Planar polygons play an important role in computer graphics. In this chapter we discover what benefits vectors bring to solving problems involving planes. We begin this exploration by examining the Cartesian form of the plane equation.

4.2 The Cartesian form of the plane equation

The general form of the plane equation creates an equation that equals zero, whereas the Cartesian form organizes the equation such that a constant term is isolated on one side of the equals sign:

$$ax + by + cz = d.$$

Both forms have their individual advantages, but the Cartesian form is useful from a geometric perspective.

We are going to use vector analysis to derive the plane equation, and the reader will see that there is an intimate relationship between this and the Cartesian form of the line equation described in Section 3.3. Furthermore, we will approach the analysis in an identical fashion.

Step 1

Define a plane away from the origin O containing the point $P(x,\ y,\ z)$ and its associated position vector $\mathbf{p}$, as shown in Fig. 4.1.

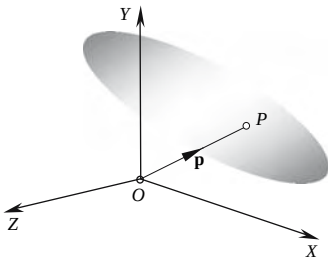


FIGURE 4.1.

Step 2

Construct a line from O to a point C on the plane such that $\overrightarrow{OC}$ is perpendicular to the plane. It is convenient to make $\overrightarrow{OC}$ a scalar multiple of a unit vector $\hat{\mathbf{n}} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, i.e., $\overrightarrow{OC} = d\hat{\mathbf{n}}$, where d is some scalar, as shown in Fig. 4.2.

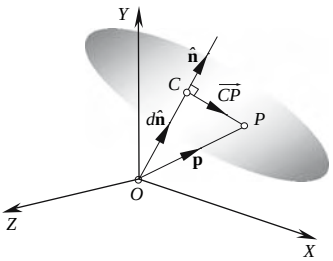


FIGURE 4.2.

We can now write $\overrightarrow{OP} = \overrightarrow{OC} + \overrightarrow{CP}$. Substituting vector names gives

$$\mathbf{p} = d\hat{\mathbf{n}} + \overrightarrow{CP}. \tag{4.1}$$

Although we know d and $\hat{\mathbf{n}}$, and the fact that $\mathbf{p}$ points to any point on the plane, we do not know $\overrightarrow{CP}$. Somehow, $\overrightarrow{CP}$ has to be eliminated, which can be achieved by the same cunning subterfuge used in Section 3.3.

Remember that the dot product of two perpendicular vectors is zero. Then $d\hat{\mathbf{n}} \cdot \overrightarrow{CP}$ must equal zero. But the d is superfluous; therefore, $\hat{\mathbf{n}} \cdot \overrightarrow{CP} = 0$.

Unfortunately, Eq. (4.1) does not contain such a term, but there is nothing to stop us from introducing one by multiplying Eq. (4.1) by $\hat{\mathbf{n}}$ using the dot product:

$$\hat{\mathbf{n}} \cdot \mathbf{p} = d\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} + \hat{\mathbf{n}} \cdot \overrightarrow{CP}$$

which now reduces to

$$\hat{\mathbf{n}} \cdot \mathbf{p} = d\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}. \tag{4.2}$$

But we already know that $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = 1$. Therefore,

$$\hat{\mathbf{n}} \cdot \mathbf{p} = d. \tag{4.3}$$

Expanding Eq. (4.3), we get

$$ax + by + cz = d$$

which we recognise as the Cartesian form of the plane equation where

x, y, z are the coordinates of a point on the plane,
$a, b,$ and c are the components of a unit vector normal to the plane, and
d is the perpendicular distance from the origin to the plane.

The above analysis assumes that the vector normal to the plane is a unit vector, which gives a precise geometric meaning to d. However, we must be careful not to apply this geometric meaning to all plane equations. For example, consider the following equation:

$$2x + 3y + 4z = 10. \tag{4.4}$$

The vector normal to the plane is clearly not a unit vector, as its magnitude is

$$\sqrt{2^2 + 3^2 + 4^2} = \sqrt{29}.$$

But if we divided Eq. (4.4) throughout by $\sqrt{29}$, we would have

$$\frac{2}{\sqrt{29}}x + \frac{3}{\sqrt{29}}y + \frac{4}{\sqrt{29}}z = \frac{10}{\sqrt{29}}.$$

The perpendicular distance from the origin to the plane is $\frac{10}{\sqrt{29}}$.

Let us now consider an alternative definition of the plane equation – the parametric form.

4.3 The parametric form of the plane equation

The parametric form of the plane equation requires two vectors and a single point to define a plane. The two vectors provide the orientation of a plane, while the single point identifies a specific plane. Such a scenario is shown in Fig. 4.3.

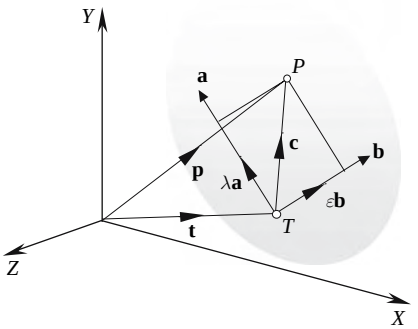


FIGURE 4.3.

Let vectors **a** and **b** be parallel to the plane and defined as

$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}$$
$$\mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}$$

and the point $T(x_T, y_T, z_T)$ is on the plane. Therefore,

$$\mathbf{c} = \lambda\mathbf{a} + \varepsilon\mathbf{b}$$

where λ and ε are scalars. But

$$\mathbf{p} = \mathbf{t} + \mathbf{c}.$$

Therefore,

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{a} + \varepsilon\mathbf{b}.$$

More explicitly,

$$x_P = x_T + \lambda x_a + \varepsilon x_b$$
$$y_P = y_T + \lambda y_a + \varepsilon y_b$$
$$z_P = z_T + \lambda z_a + \varepsilon z_b.$$

If **a** and **b** are unit vectors and are mutually perpendicular, i.e., $\mathbf{a} \cdot \mathbf{b} = 0$, λ and ε become linear measurements along the **a**- and **b**-axes relative to T.

4.4 A plane equation from three points

We know that two points are required to define a line, but three points are needed to define a plane. So let us now consider how the plane equation is derived from three such points.

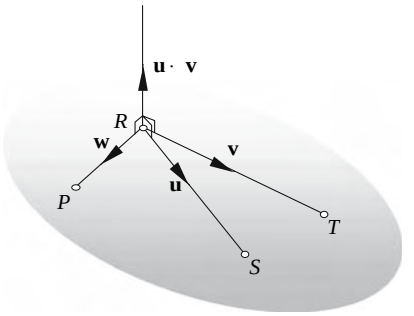


FIGURE 4.4.

We start with three points, $R(x_R, y_R, z_R)$, $S(x_S, y_S, z_S)$, and $T(x_T, y_T, z_T)$, as shown in Fig. 4.4, and create two vectors, $\mathbf{u} = \overrightarrow{RS}$ and $\mathbf{v} = \overrightarrow{RT}$. The three points are assumed to be in a counterclockwise sequence viewed from the direction of the surface normal. The vector product $\mathbf{u} \times \mathbf{v}$ provides a vector normal to the plane containing the points:

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix}.$$

We now take another point, $P(x, y, z)$, which is on the plane, and form a vector, $\mathbf{w} = \overrightarrow{RP}$. Therefore, the scalar product $\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = 0$. This condition can be expressed as a determinant and converted into the general equation of a plane:

$$\mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) = \begin{vmatrix} x_w & y_w & z_w \\ x_u & y_u & z_u \\ x_v & y_v & z_v \end{vmatrix} = 0. \tag{4.5}$$

Expanding Eq. (4.5), we obtain

$$x_w \begin{vmatrix} y_u & z_u \\ y_v & z_v \end{vmatrix} + y_w \begin{vmatrix} z_u & x_u \\ z_v & x_v \end{vmatrix} + z_w \begin{vmatrix} x_u & y_u \\ x_v & y_v \end{vmatrix} = 0$$

which becomes

$$(x - x_R) \begin{vmatrix} y_S - y_R & z_S - z_R \\ y_T - y_R & z_T - z_R \end{vmatrix} + (y - y_R) \begin{vmatrix} z_S - z_R & x_S - x_R \\ z_T - z_R & x_T - x_R \end{vmatrix} + (z - z_R) \begin{vmatrix} x_S - x_R & y_S - y_R \\ x_T - x_R & y_T - y_R \end{vmatrix} = 0.$$

This can be arranged in the Cartesian form of the line equation

$$ax + by + cz = d$$

where

$$a = \begin{vmatrix} y_S - y_R & z_S - z_R \\ y_T - y_R & z_T - z_R \end{vmatrix} \quad \text{or} \quad a = \begin{vmatrix} 1 & y_R & z_R \\ 1 & y_S & z_S \\ 1 & y_T & z_T \end{vmatrix}$$

$$b = \begin{vmatrix} z_S - z_R & x_S - x_R \\ z_T - z_R & x_T - x_R \end{vmatrix} \quad \text{or} \quad b = \begin{vmatrix} x_R & 1 & z_R \\ x_S & 1 & z_S \\ x_T & 1 & z_T \end{vmatrix}$$

$$c = \begin{vmatrix} x_S - x_R & y_S - y_R \\ x_T - x_R & y_T - y_R \end{vmatrix} \quad \text{or} \quad c = \begin{vmatrix} x_R & y_R & 1 \\ x_S & y_S & 1 \\ x_T & y_T & 1 \end{vmatrix}$$

$$d = ax_R + by_R + cz_R.$$

4.5 A plane perpendicular to a line and passing through a point

Although there are an infinite number of planes perpendicular to a given line, simply by defining a point that lies on a plane enables us to fix a unique plane. In this section we derive a formula that defines the Cartesian form of the plane equation using a line and a point. The line represents a vector perpendicular to a plane, while the point identifies a specific plane in space.

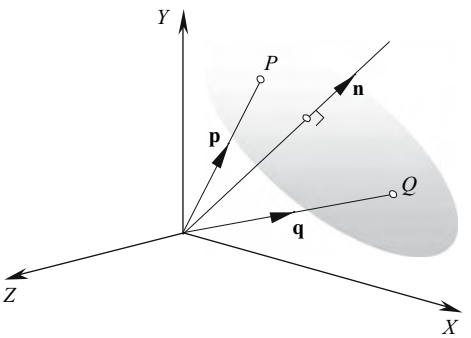


Figure 4.5.

With reference to Fig. 4.5, let the plane equation be

$$ax + by + cz = d$$

and the line direction vector be

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

and the specified point on the plane be

$$Q\left(x_Q, y_Q, z_Q\right).$$

If P is any point on the plane with position vector

$$\mathbf{p} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

then

$$\mathbf{n} \cdot \mathbf{p} = d. \tag{4.6}$$

As $\mathbf{p}$ and $\mathbf{q}$ share a common projection on $\mathbf{n}$,

$$\mathbf{n} \cdot \mathbf{p} = \mathbf{n} \cdot \mathbf{q}. \tag{4.7}$$

Substituting Eq. (4.6) in Eq. (4.7) gives

$$\mathbf{n} \cdot \mathbf{q} = d$$

$$ax_Q + by_Q + cz_Q = d. \tag{4.8}$$

Substituting Eq. (4.8) in Eq. (4.6) gives

$$ax + by + cz = ax_Q + by_Q + cz_Q. \tag{4.9}$$

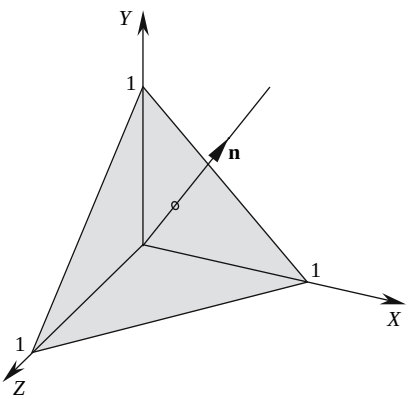


Figure 4.6.

To test Eq. (4.9), consider the scenario shown in Fig. 4.6. The line's direction vector $\mathbf{n}$ is given by

$$\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$$

and the point on the plane is $Q(1, 0, 0)$. Therefore, the plane's equation is

$$ax + by + cz = 1.$$

4.6 A plane through two points and parallel to a line

The task in this section is to create a Cartesian plane equation from two points and a line. The line is parallel with the plane, rather than being perpendicular to the plane, and the two points are used to locate a second line and provide a point that resides on the plane. We begin by creating two vectors: one from the two points and the other from the line. The vector-product of these vectors is normal to the associated plane. Figure 4.7 shows the two points M and N, and the direction vector $\lambda\mathbf{a}$ representing the line.

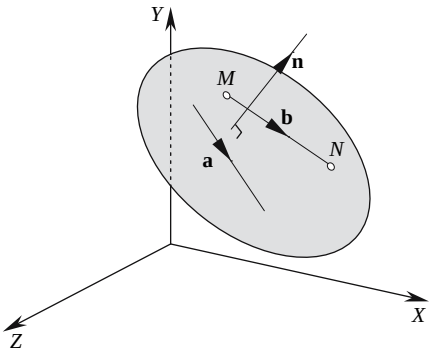


FIGURE 4.7.

Let the line be defined by

$$\mathbf{p} = \mathbf{r} + \lambda\mathbf{a}$$

where

$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}.$$

The two points are

$$M\,(x_M, y_M, z_M) \text{ and } N\,(x_N, y_N, z_N)\,.$$

Therefore, the vector $\mathbf{b}$ is given by

$$\mathbf{b} = (x_N - x_M)\,\mathbf{i} + (y_N - y_M)\,\mathbf{j} + (z_N - z_M)\,\mathbf{k}.$$

The vector $\mathbf{n}$ perpendicular to the plane is the cross product of $\mathbf{a}$ and $\mathbf{b}$:

$$\mathbf{a} \times \mathbf{b} = \mathbf{n}$$

where

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

and

$$a = \begin{vmatrix} y_a & z_a \\ y_b & z_b \end{vmatrix} \quad b = \begin{vmatrix} z_a & x_a \\ z_b & x_b \end{vmatrix} \quad c = \begin{vmatrix} x_a & y_a \\ x_b & y_b \end{vmatrix}.$$

Therefore, the plane equation is

$$ax + by + cz = d.$$

To find d, we chose one of the two points on the plane, say M, and we can state

$$ax_M + by_M + cz_M = d.$$

Therefore,

$$ax + by + cz = ax_M + by_M + cz_M. \tag{4.10}$$

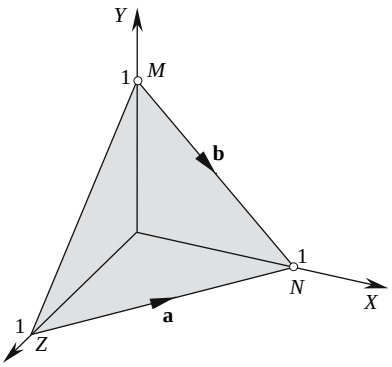


Figure 4.8.

We can test Eq. (4.10) using the same example in the previous section, as shown in Fig. 4.8. The line's direction vector **a** is given by

$$\mathbf{a} = \mathbf{i} - \mathbf{k}$$

and the two points M and N are

$$M\,(0, 1, 0) \text{ and } N\,(1, 0, 0)\,.$$

Therefore,

$$a = \begin{vmatrix} 0 & -1 \\ -1 & 0 \end{vmatrix} \quad b = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} \quad c = \begin{vmatrix} 1 & 0 \\ 1 & -1 \end{vmatrix}$$

$$a = -1 \quad b = -1 \quad c = -1.$$

If we evaluate Eq. (4.10) using a, b, c, and M, we obtain

$$-x-y-z=-1. \tag{4.11}$$

We can tidy up Eq. (4.11) by multiplying throughout by –1 to give

$$x+y+z=1.$$

Now let's examine the ideas behind 3D space partitioning.

4.7 3D space partitioning

In Section 3.5 we discovered how it was possible to determine whether a point was inside or outside a convex boundary. The technique also revealed whether a point was on a boundary vertex or edge. Now let's explore a similar technique where we can determine whether a point is inside, outside, or on a vertex, edge, or surface forming a convex polyhedron.

A simple example reveals the underlying mechanism of this form of 3D space partitioning. A unit cube has been chosen aligned with the x-, y-, and z-axes, as it keeps the mathematics very simple. However, it should be obvious from this example that the mathematical reasoning holds for any convex polyhedron.

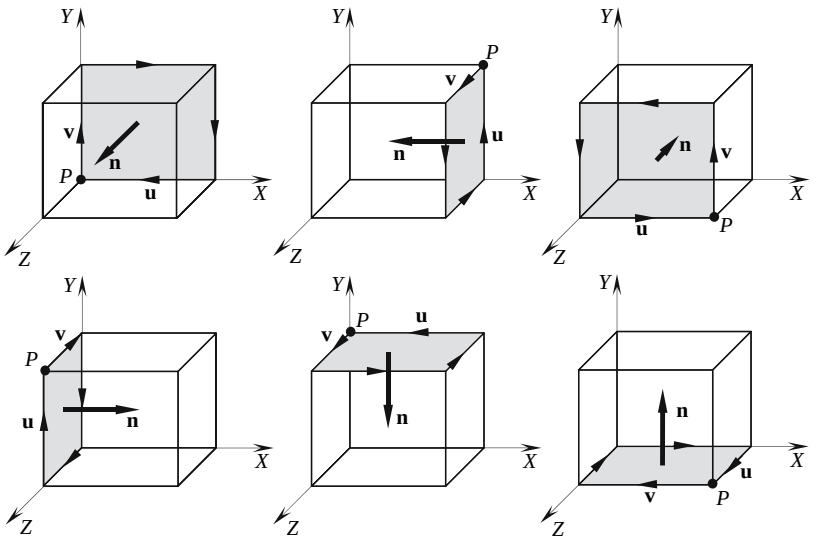


FIGURE 4.9.

Figure 4.9 shows six views of a unit cube with each surface highlighted with its normal vector **n**. To ensure that all the normal vectors point inwards, it is derived from two similar edges forming each surface. We assume that the surfaces are defined by a chain of edges defined in a counter-clockwise sequence when viewed from the outside, as shown in Fig. 4.9.

The cross-product operation is used to create a normal vector from two adjacent edges selected from each surface. Any two will do, but the selection procedure must be the same for all the surfaces, otherwise the normal vector will not point consistently inwards or outwards. The chosen edges are the vectors **u** and **v**, as shown in Fig. 4.9.

By reversing one of them, say **u**, we can see that

$$\mathbf{n} = -\mathbf{u} \times \mathbf{v}.$$

Furthermore, given a vector

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

the plane equation is given by

$$ax + by + cz = ax_P + by_P + cz_P$$

where (x_P, y_P, z_P) is a point on the plane. Such a point is shown on each view in Fig. 4.9.

For example, in the first view,

$$\mathbf{u} = -\mathbf{i} \quad \mathbf{v} = \mathbf{j} \quad \mathbf{n} = \mathbf{k} \quad P = (0, 0, 0)$$

therefore, the plane equation is given by $z = 0$. These results are shown in the first row of Table 4.1, followed by the other five surfaces.

The LHS of the plane equation returns a value of zero for any point on the plane; a positive or negative value is returned for points not on the plane. If the point is in the space partition occupied by the surface normal, a positive value is returned, otherwise it is negative.

Table 4.2 illustrates how the LHS expressions of the plane equations react to five different points. With reference to Fig. 4.10, the point $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ is obviously inside the convex volume, and all the expressions are positive, whereas the point $(1, 0, 2)$ is outside the volume and the expression $-z + 1$ returns a negative value. Consequently, if any expression goes negative, the search can halt and the point is declared outside.

If a point resides on one of the surfaces, the corresponding expression returns a zero value, as shown with the point $\left(\frac{1}{2}, 1, \frac{1}{2}\right)$. If a point resides on an edge, two expressions return a zero value, as shown with the point $\left(\frac{1}{2}, 0, 1\right)$. Finally, if a point resides on a vertex, three expressions return a zero value, as shown with the point $(1, 1, 1)$. The table entries that determine the final result are shaded.

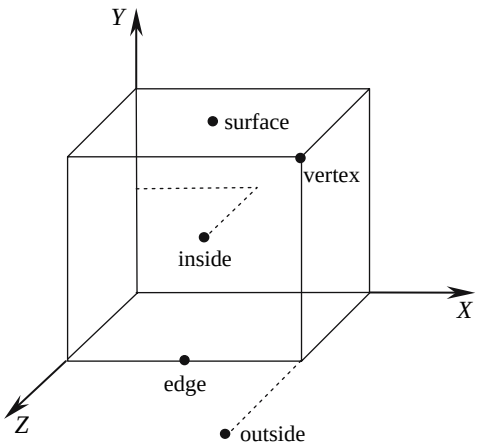


Figure 4.10.

Table 4.1

Surface	**u**	**v**	**n** $=-\mathbf{u}\times\mathbf{v}$	P	Equation
1	$-\mathbf{i}$	**j**	**k**	$(0,0,0)$	$z=0$
2	**j**	**k**	$-\mathbf{i}$	$(1,1,0)$	$-x+1=0$
3	**i**	**j**	$-\mathbf{k}$	$(1,0,1)$	$-z+1=0$
4	**j**	$-\mathbf{k}$	**i**	$(0,1,1)$	$x=0$
5	$-\mathbf{i}$	**k**	$-\mathbf{j}$	$(0,1,0)$	$-y+1=0$
6	**k**	$-\mathbf{i}$	**j**	$(1,0,1)$	$y=0$

Table 4.2

LHS expression	$\left(\frac{1}{2},\frac{1}{2},\frac{1}{2}\right)$	$(1,0,2)$	$\left(\frac{1}{2},1,\frac{1}{2}\right)$	$\left(\frac{1}{2},0,1\right)$	$(1,1,1)$
z	+	+	+	+	+
$-x+1$	+	0	+	+	0
$-z+1$	+	–	+	0	0
x	+	+	+	+	+
$-y+1$	+	+	0	+	0
y	+	0	+	0	+
Result	**inside**	**outside**	**surface**	**edge**	**vertex**

Hopefully, this simple exercise has revealed just how useful vectors are in resolving a very useful space-partitioning technique.

4.8 The angle between two planes

The angle between two planes is readily computed using the dot product of the two surface normals. For example, given two Cartesian plane equations

$$ax_1 + by_1 + cz_1 = d_1$$

and

$$ax_2 + by_2 + cz_2 = d_2$$

the normal vectors are

$$\mathbf{n}_1 = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$$

and

$$\mathbf{n}_2 = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}.$$

Therefore,

$$\mathbf{n}_1 \cdot \mathbf{n}_2 = \|\mathbf{n}_1\| \, \|\mathbf{n}_2\| \cos\alpha$$

and

$$\alpha = \cos^{-1}\left(\frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{\|\mathbf{n}_1\| \, \|\mathbf{n}_2\|}\right).$$

If

$$\|\mathbf{n}_1\| = \|\mathbf{n}_2\| = 1$$

then

$$\alpha = \cos^{-1}(\mathbf{n}_1 \cdot \mathbf{n}_2).$$

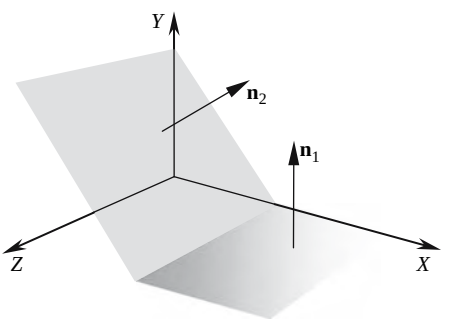


Figure 4.11.

For example, Fig. 4.11 shows two planes whose plane equations are given by

$$y = 0$$

and

$$x + y = 1.$$

Therefore, the two normal vectors are

$$\mathbf{n}_1 = \mathbf{j}$$

and

$$\mathbf{n}_2 = \mathbf{i} + \mathbf{j}.$$

Therefore,

$$\alpha = \cos^{-1}\left(\frac{\mathbf{j}\cdot(\mathbf{i}+\mathbf{j})}{1\times\sqrt{2}}\right) = \cos^{-1}\left(\frac{\sqrt{2}}{2}\right) = 45°.$$

4.9 The angle between a line and a plane

The angle between a line and a plane is virtually identical to that used to find the angle between two planes. All that is required is the line's direction vector and plane's normal vector. The dot product of the two vectors will then reveal their separating angle.

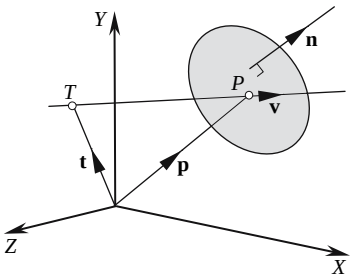


Figure 4.12.

For example, Fig. 4.12 shows a line and a plane where the line is defined as

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

and the plane equation by

$$ax + by + cz = d.$$

The plane's surface normal is

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Therefore,

$$\mathbf{n} \cdot \mathbf{v} = \|\mathbf{n}\| \cdot \|\mathbf{v}\| \cos\alpha$$

and

$$\alpha = \cos^{-1}\left(\frac{\mathbf{n} \cdot \mathbf{v}}{\|\mathbf{n}\| \cdot \|\mathbf{v}\|}\right)$$

If

$$\|\mathbf{n}\| = \|\mathbf{v}\| = 1$$

then

$$\alpha = \cos^{-1}(\mathbf{n} \cdot \mathbf{v}).$$

when the line is parallel with the plane $\mathbf{n} \cdot \mathbf{v} = 0$

4.10 The position and distance of the nearest point on a plane to a point

This problem is concerned with the relationship between a point and a plane; in particular, the problem of finding the position and distance of a point on a plane that is the nearest to some specific point. For example, Fig. 4.13 shows a polygon representing a ground plane and a point P in space. The problem is to find the location and distance of a point on the plane that is nearest to P. Hopefully, it is obvious that the nearest point on the plane is perpendicular to P, which means that the dot product plays some role in the solution. Let's examine how vector analysis reveals a solution.

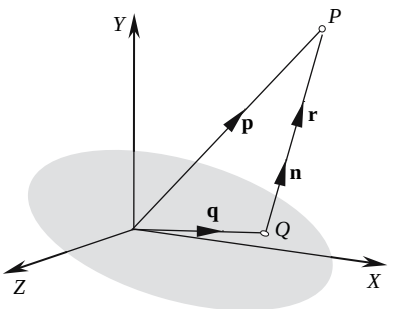


Figure 4.13.

Let the plane equation be

$$ax + by + cz = d.$$

Therefore, the surface normal $\mathbf{n}$ is given by

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Let $Q(x, y, z)$ be the nearest point on the plane to P. Therefore,

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

and

$$\mathbf{n} \cdot \mathbf{q} = d. \tag{4.12}$$

Let

$$\overrightarrow{QP} = \mathbf{r} = \lambda\mathbf{n}.$$

Therefore,

$$\mathbf{q} = \mathbf{p} - \lambda\mathbf{n}. \tag{4.13}$$

To find $\mathbf{q}$, we need to find λ.

Using $\mathbf{n}$, take the dot product of Eq. (4.13):

$$\mathbf{n} \cdot \mathbf{q} = \mathbf{n} \cdot \mathbf{p} - \lambda\mathbf{n} \cdot \mathbf{n} \tag{4.14}$$

Substitute Eq. (4.12) in Eq. (4.14):

$$d = \mathbf{n} \cdot \mathbf{p} - \lambda\mathbf{n} \cdot \mathbf{n}$$

and

$$\lambda = \frac{\mathbf{n} \cdot \mathbf{p} - d}{\mathbf{n} \cdot \mathbf{n}} = \frac{\mathbf{n} \cdot \mathbf{p} - d}{\|\mathbf{n}\|^2}.$$

If

$$\|\mathbf{n}\| = 1$$

then

$$\lambda = \mathbf{n} \cdot \mathbf{p} - d.$$

Q is given by

$$\mathbf{q} = \mathbf{p} - \lambda\mathbf{n}. \tag{4.15}$$

The distance of Q is

$$PQ = \|\lambda \mathbf{n}\|.$$

If

$$\|\mathbf{n}\| = 1$$

then

$$PQ = \|\lambda\|.$$

Let's test Eq. (4.15) with an example.

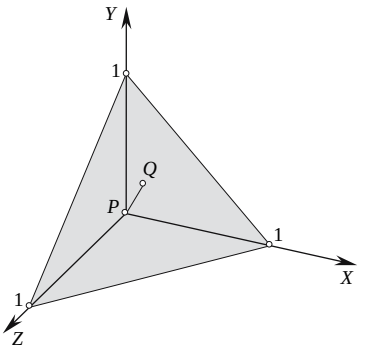


FIGURE 4.14.

Figure 4.14 shows a triangle whose plane equation is

$$x + y + z = 1.$$

Therefore,

$$\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$$

and

$$d = 1.$$

We will find the nearest point on the triangle to the origin by making $P = (0, 0, 0)$.

Therefore,

$$\lambda = \frac{\mathbf{n} \cdot \mathbf{p} - d}{\|\mathbf{n}\|^2} = \frac{(\mathbf{i} + \mathbf{j} + \mathbf{k}) \cdot (0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}) - 1}{\sqrt{3}\sqrt{3}} = -\frac{1}{3}$$

and

$$\mathbf{q} = \mathbf{p} - \lambda \mathbf{n} = (0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}) - \left(-\frac{1}{3}\right)(\mathbf{i} + \mathbf{j} + \mathbf{k})$$

$$\mathbf{q} = \frac{1}{3}\mathbf{i} + \frac{1}{3}\mathbf{j} + \frac{1}{3}\mathbf{k}$$

Therefore, the nearest point to P is $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$.

4.11 The reflection of a point in a plane

When dealing with mirrors, we need to know how to compute the virtual position of an object's reflection. Once we know how to compute a single point, most objects can be processed. Central to the solution of this problem is the fact that the virtual reflection is the same distance behind the mirror as the object is in front of the mirror. Furthermore, a line connecting a real point to its virtual reflection is orthogonal to the mirror (see Fig. 4.15).

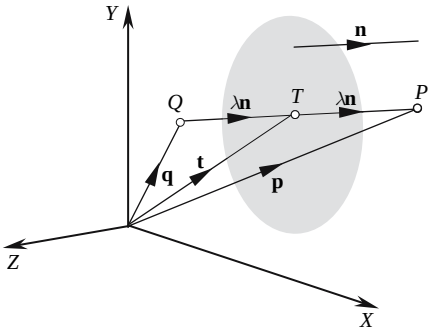


FIGURE 4.15.

Let the mirror's plane equation be

$$ax + by + cz = d$$

then the vector normal is

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Let T be a point on the mirror such that $\overrightarrow{TP}$ is parallel to $\mathbf{n}$. Let T's position vector be

$$\mathbf{t} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

Therefore,

$$\mathbf{n} \cdot \mathbf{t} = ax + by + cz = d. \tag{4.16}$$

Now

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{n} \tag{4.17}$$

where λ is a scalar.

Using $\mathbf{n}$, take the dot product of Eq. (4.17):

$$\mathbf{n} \cdot \mathbf{p} = \mathbf{n} \cdot \mathbf{t} + \lambda\mathbf{n} \cdot \mathbf{n} \tag{4.18}$$

Substituting Eq. (4.16) in Eq. (4.18) gives

$$\mathbf{n}\cdot\mathbf{p} = d + \lambda\mathbf{n}\cdot\mathbf{n}.$$

Therefore,

$$\lambda = \frac{\mathbf{n}\cdot\mathbf{p} - d}{\mathbf{n}\cdot\mathbf{n}} = \frac{\mathbf{n}\cdot\mathbf{p} - d}{\|\mathbf{n}\|^2}. \tag{4.19}$$

But

$$\mathbf{q} = \mathbf{p} - 2\lambda\mathbf{n}.$$

Therefore,

$$\mathbf{q} = \mathbf{p} - 2\left(\frac{\mathbf{n}\cdot\mathbf{p} - d}{\|\mathbf{n}\|^2}\right)\mathbf{n} \tag{4.20}$$

which enables us calculate the location of Q.

If $\mathbf{n}$ is a unit vector $\hat{\mathbf{n}}$, then

$$\mathbf{q} = \mathbf{p} - 2\,(\hat{\mathbf{n}}\cdot\mathbf{p} - d)\,\hat{\mathbf{n}}. \tag{4.21}$$

Let's explore two scenarios for these equations: one with a point behind the mirror and one in front, which will demonstrate that mathematical mirrors possess two reflecting sides.

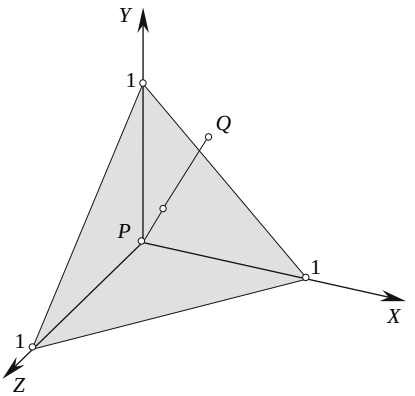


Figure 4.16.

Figure 4.16 shows a triangular mirror whose plane equation is given by

$$x + y + z = 1$$

where

$$\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$$

and

$$d = 1.$$

To start with, let

$$P = (0, 0, 0).$$

Using Eq. (4.20) gives

$$\mathbf{q} = (0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}) - 2\left(\frac{(\mathbf{i}+\mathbf{j}+\mathbf{k}) \cdot (0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}) - 1}{\sqrt{3}\sqrt{3}}\right)(\mathbf{i}+\mathbf{j}+\mathbf{k})$$

$$\mathbf{q} = \frac{2}{3}\mathbf{i} + \frac{2}{3}\mathbf{j} + \frac{2}{3}\mathbf{k}.$$

Therefore, the reflection is in the space partition occupied by the mirror's normal $\left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$.

If we now place P at $\left(\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$, its reflection should be at the origin.

Using Eq. (4.20) gives

$$\mathbf{q} = \left(\frac{2}{3}\mathbf{i} + \frac{2}{3}\mathbf{j} + \frac{2}{3}\mathbf{k}\right) - 2\left(\frac{(\mathbf{i}+\mathbf{j}+\mathbf{k}) \cdot \left(\frac{2}{3}\mathbf{i} + \frac{2}{3}\mathbf{j} + \frac{2}{3}\mathbf{k}\right) - 1}{\sqrt{3}\sqrt{3}}\right)(\mathbf{i}+\mathbf{j}+\mathbf{k})$$

$$\mathbf{q} = \left(\frac{2}{3}\mathbf{i} + \frac{2}{3}\mathbf{j} + \frac{2}{3}\mathbf{k}\right) - \frac{2}{3}(\mathbf{i}+\mathbf{j}+\mathbf{k}) = 0\mathbf{i} + 0\mathbf{j} + 0\mathbf{k}$$

making the reflection at the origin.

4.12 A plane between two points

Given two distinct points in $\mathbb{R}^3$, we can connect a straight line between them. Furthermore, we can locate a plane orthogonal to this line and positioned such that it divides the line into the ratio $t : (1 - t)$, where $t \in [0, 1]$. Let's examine how such a plane is identified.

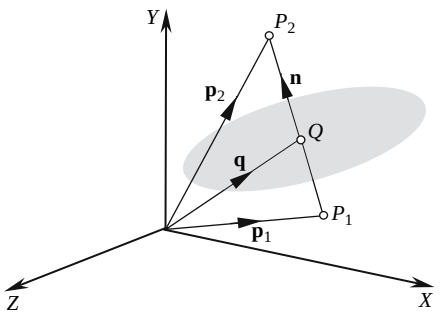


Figure 4.17.

Figure 4.17 shows two points, $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$, joined by a connecting line. Let the plane equation be of the form

$$ax + by + cz = d.$$

Let the plane's normal $\mathbf{n}$ be

$$\mathbf{n} = \mathbf{p}_2 - \mathbf{p}_1 \tag{4.22}$$

and

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Let $Q(x, y, z)$ be a point on $\overrightarrow{P_1P_2}$ such that

$$\mathbf{q} = (1-t)\,\mathbf{p}_1 + t\mathbf{p}_2. \tag{4.23}$$

Therefore,

$$\mathbf{n} \cdot \mathbf{q} = d. \tag{4.24}$$

Substituting Eq. (4.22) and Eq. (4.23) in Eq. (4.24) gives

$$d = (\mathbf{p}_2 - \mathbf{p}_1) \cdot ((1-t)\,\mathbf{p}_1 + t\mathbf{p}_2).$$

Finally, we have

$$a = x_2 - x_1 \qquad b = y_2 - y_1 \qquad c = z_2 - z_1$$
$$d = a((1-t)\,x_1 + tx_2) + b((1-t)\,y_1 + ty_2) + c((1-t)\,z_1 + tz_2).$$

Let's test the above equations with the scenario shown in Fig. 4.18.

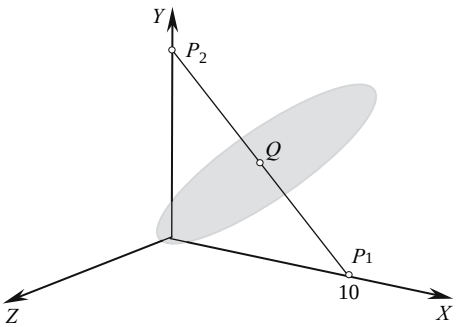


FIGURE 4.18.

Given $P_1(10, 0, 0)$, $P_2(0, 10, 0)$, and $t = \frac{1}{2}$, we have

$$a = x_2 - x_1 = -10 \qquad b = y_2 - y_1 = 10 \qquad c = z_2 - z_1 = 0$$
$$d = -10(5) + 10(5) = 0.$$

Therefore, the plane equation is

$$-10x + 10y = 0.$$

5 Reflections

5.1 Introduction

In this chapter we explore two scenarios where lines reflect from lines and surfaces. The reflecting lines can represent the path taken by a collection of photons or the trajectory of a small particle when striking another line or surface. Once more we will discover that the vector equations describing a 2D scenario apply equally well in a 3D scenario. We begin by calculating how a 2D ray bounces from a reflecting line.

5.2 A line reflecting off a line

The reflecting line is defined parametrically, and a second parametric line represents the incoming ray striking the reflecting line. At the point of contact, reflection takes place obeying the classical laws of reflection: i.e., the angle of reflection equals the angle of incidence, and the normal, incident, and reflected rays lie in a common plane. Figure 5.1 illustrates this scenario.

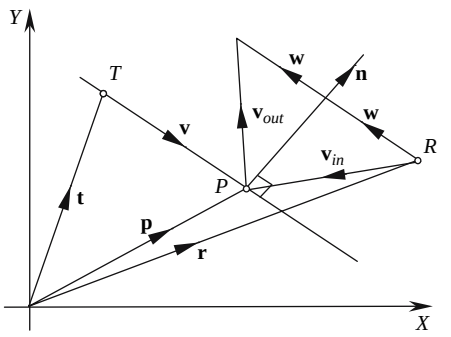


Figure 5.1.

Let the reflecting line be defined as

$$\mathbf{p} = \mathbf{t} + \varepsilon \mathbf{v} \tag{5.1}$$

where ε is a scalar, and the line's normal vector is

$$\mathbf{n} = \mathbf{v}^{\perp}.$$

The incoming ray passes through R with direction $\mathbf{v}_{\text{in}}$, and the reflected outgoing ray originates at P with direction $\mathbf{v}_{\text{out}}$.

From Fig. 5.1 we observe that

$$\mathbf{w} = \mathbf{v}_{\text{in}} + \lambda \mathbf{n} \tag{5.2}$$

and

$$\mathbf{w} = \mathbf{v}_{\text{out}} - \lambda \mathbf{n} \tag{5.3}$$

where λ is a scalar.

Equating (5.2) and (5.3) gives

$$\mathbf{v}_{\text{in}} + \lambda \mathbf{n} = \mathbf{v}_{\text{out}} - \lambda \mathbf{n}$$

and

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} + 2\lambda \mathbf{n}. \tag{5.4}$$

We now need to find λ.

We multiply Eq. (5.4) throughout by $\mathbf{n}$ and obtain

$$\mathbf{n} \cdot \mathbf{v}_{\text{out}} = \mathbf{n} \cdot \mathbf{v}_{\text{in}} + 2\lambda \mathbf{n} \cdot \mathbf{n}$$

but

$$\mathbf{n} \cdot \mathbf{v}_{\text{out}} = \mathbf{n} \cdot (-\mathbf{v}_{\text{in}}) = -\mathbf{n} \cdot \mathbf{v}_{\text{in}}.$$

Therefore,

$$-\mathbf{n} \cdot \mathbf{v}_{\text{in}} = \mathbf{n} \cdot \mathbf{v}_{\text{in}} + 2\lambda \mathbf{n} \cdot \mathbf{n}$$

and

$$\lambda = -\frac{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}{\| \mathbf{n} \|^2}. \tag{5.5}$$

Substituting Eq. (5.5) in Eq. (5.4) gives

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2\left(\frac{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}{\|\mathbf{n}\|^2}\right)\mathbf{n}. \tag{5.6}$$

Note that if $\mathbf{n}$ is a unit vector, written as $\hat{\mathbf{n}}$, then

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2\,(\hat{\mathbf{n}} \cdot \mathbf{v}_{\text{in}})\,\hat{\mathbf{n}}.$$

We now need to find P and its position vector $\mathbf{p}$. From Fig. 5.1, we see that

$$\mathbf{p} = \mathbf{r} + \kappa \mathbf{v}_{\text{in}} \tag{5.7}$$

where κ is a scalar. Equating (5.1) and (5.7) gives

$$\mathbf{t} + \varepsilon \mathbf{v} = \mathbf{r} + \kappa \mathbf{v}_{\text{in}}. \tag{5.8}$$

We multiply Eq. (5.8) throughout by $\mathbf{n}$ and obtain

$$\mathbf{n} \cdot \mathbf{t} + \varepsilon \mathbf{n} \cdot \mathbf{v} = \mathbf{n} \cdot \mathbf{r} + \kappa \mathbf{n} \cdot \mathbf{v}_{\text{in}}$$

but

$$\mathbf{n} \cdot \mathbf{v} = 0.$$

Therefore,

$$\mathbf{n} \cdot \mathbf{t} - \mathbf{n} \cdot \mathbf{r} = \kappa \mathbf{n} \cdot \mathbf{v}_{\text{in}}$$

and

$$\kappa = \frac{\mathbf{n} \cdot (\mathbf{t} - \mathbf{r})}{\mathbf{n} \cdot \mathbf{v}_{\text{in}}} \tag{5.9}$$

Substituting Eq. (5.9) in Eq. (5.7) gives

$$\mathbf{p} = \mathbf{r} + \left(\frac{\mathbf{n} \cdot (\mathbf{t} - \mathbf{r})}{\mathbf{n} \cdot \mathbf{v}_{\text{in}}} \right) \mathbf{v}_{\text{in}}. \tag{5.10}$$

Therefore, the reflected ray originates at P with direction $\mathbf{v}_{\text{out}}$.

Let's test these results with an example. Figure 5.2 shows a simple scenario that permits us to predict the outcome. As the reflecting line makes an angle of 45° with the x-axis, and $\mathbf{v}_{\text{in}}$ is parallel with the x-axis, the reflected ray will be vertical, and the point of reflection is $(1, 1)$.

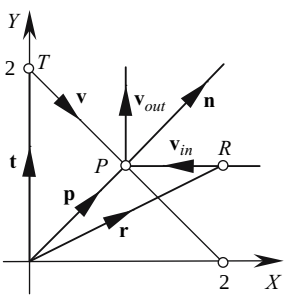


Figure 5.2.

Given the following conditions:

$$\mathbf{v} = \mathbf{i} - \mathbf{j}$$
$$R = (2, 1) \quad \text{and} \quad \mathbf{r} = 2\mathbf{i} + \mathbf{j}$$

we find that

$$\mathbf{v}_{\text{in}} = -\mathbf{i}.$$

Therefore,

$$\mathbf{n} = \mathbf{i} + \mathbf{j}.$$

Using Eq. (5.6) gives

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2\left(\frac{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}{\|\mathbf{n}\|^2}\right)\mathbf{n} = (-\mathbf{i}) - 2\left(\frac{(\mathbf{i}+\mathbf{j}) \cdot (-\mathbf{i})}{\sqrt{2}^2}\right)(\mathbf{i}+\mathbf{j})$$
$$\mathbf{v}_{\text{out}} = (-\mathbf{i}) - 2\left(\frac{-1}{2}\right)(\mathbf{i}+\mathbf{j})$$
$$\mathbf{v}_{\text{out}} = \mathbf{j}.$$

Using Eq. (5.10) gives

$$\mathbf{p} = \mathbf{r} + \left(\frac{\mathbf{n} \cdot (\mathbf{t} - \mathbf{r})}{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}\right)\mathbf{v}_{\text{in}} = (2\mathbf{i} + \mathbf{j}) + \left(\frac{(\mathbf{i}+\mathbf{j}) \cdot (2\mathbf{j} - 2\mathbf{i} - \mathbf{j})}{(\mathbf{i}+\mathbf{j}) \cdot (-\mathbf{i})}\right)(-\mathbf{i})$$
$$\mathbf{p} = (2\mathbf{i} + \mathbf{j}) + \left(\frac{(\mathbf{i}+\mathbf{j}) \cdot (-2\mathbf{i} + \mathbf{j})}{-1}\right)(-\mathbf{i}) = (2\mathbf{i} + \mathbf{j}) - \mathbf{i}$$
$$\mathbf{p} = \mathbf{i} + \mathbf{j}.$$

Therefore, P has coordinates $(1, 1)$, which confirms our prediction.

5.3 A line reflecting off a plane

You have probably already guessed that the above vector equations will work just as well in 3D as they do in 2D. Therefore, they will describe how a ray will bounce off a surface. To compute the direction of the reflected vector, all we need is the plane's surface normal and the incoming vector, which means that the equation describing $\mathbf{v}_{\text{out}}$ still holds:

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2\,(\hat{\mathbf{n}} \cdot \mathbf{v}_{\text{in}})\,\hat{\mathbf{n}}$$

and the position vector for P is

$$\mathbf{p} = \mathbf{r} + \left(\frac{\mathbf{n} \cdot (\mathbf{t} - \mathbf{r})}{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}\right)\mathbf{v}_{\text{in}}.$$

However, in lighting calculations, the vector $\mathbf{v}_{\text{in}}$ is often reversed and points toward the light, which means that its sign must be reversed in these equations:

$$\mathbf{v}_{\text{out}} = 2\,(\hat{\mathbf{n}} \cdot \mathbf{v}_{\text{in}})\,\hat{\mathbf{n}} - \mathbf{v}_{\text{in}} \tag{5.11}$$

$$\mathbf{p} = \mathbf{r} + \left(\frac{\mathbf{n} \cdot (\mathbf{t} - \mathbf{r})}{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}\right) \mathbf{v}_{\text{in}}.$$

We use Eq. (5.11) later in Section 10.2.

6 Intersections

6.1 Introduction

In this chapter we explore how vector-based techniques are used to resolve a variety of problems that arise in computer graphics such as proximity testing, modeling, collision detection, ray casting, and ray tracing.

In some cases, as we shall see, we can work directly with vector-based equations without incorporating extra transformations. For example, to compute the intersection of two lines, we can work directly with the line equations. But say we wish to compute the intersection of a cylinder with a line? Then it is highly likely that the cylinder will have to be scaled, rotated, and translated away from the origin, in which case the cylinder's equations have to be transformed before performing the intersection. Unfortunately, the resulting solution is neither simple nor elegant. However, it just so happens that if we transform the line instead such that the relative orientation between the line and cylinder is preserved, the geometric analysis remains both simple and elegant.

For instance, let's assume that an object O is modeled such that its center is at the origin, and then O is subjected to three transforms: **S** (scale), **R** (rotate), and **T** (translate) to produce the transformed object O':

$$O' = \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{S} \cdot O.$$

If we attempt to compute an intersection with O' using a ray described by

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}$$

the algebraic analysis becomes rather convoluted. However, if we leave the object where it is and transform the ray instead, we just have to make sure that the ray is transformed such that the relative orientation between the ray and the object is preserved. Basically, this requires finding the inverse transforms of **T**, **R**, and **S**:

$$\mathbf{T}^{-1}, \ \mathbf{R}^{-1} \text{ and } \mathbf{S}^{-1}$$

and applying them to the ray. But before so doing, we must remember that vectors are rather sensitive to transforms.

For example, the above line equation contains two types of vector: $\mathbf{t}$ is a position vector with its tail fixed at the origin and its head locating some point on the ray, typically T; while $\mathbf{v}$ is the ray's direction in terms of Cartesian displacements. Both types of vector behave correctly when scaled and rotated, but only $\mathbf{t}$ can be translated. Therefore, the ray must be transformed as follows:

$$\mathbf{t}' = \mathbf{S}^{-1} \cdot \mathbf{R}^{-1} \cdot \mathbf{T}^{-1} \cdot \mathbf{t}$$

and

$$\mathbf{v}' = \mathbf{S}^{-1} \cdot \mathbf{R}^{-1} \cdot \mathbf{v}.$$

The transformed ray becomes

$$\mathbf{p} = \mathbf{t}' + \lambda \mathbf{v}'.$$

with which we can test for a possible intersection with the untransformed object. So, with this in mind, some of the following analyses will only consider objects centered about the origin, and the relative juxtaposition between ray and object is achieved by transforming the ray.

No attempt is made in the following sections to develop fast algorithms. The underlying goals are to identify strategies and to reveal patterns in solutions.

Let's begin this survey with the simple case of two intersecting lines in $\mathbb{R}^2$.

6.2 Two intersecting lines in $\mathbb{R}^2$

Given two lines in $\mathbb{R}^2$, there are three possible orientation scenarios: they intersect, they are parallel, or they are identical. We begin by assuming that they intersect, as shown in Fig. 6.1. The lines are defined using a direction vector and a known point on the line. The point of intersection P must satisfy both equations, and our task is to compute the value of λ and ε.

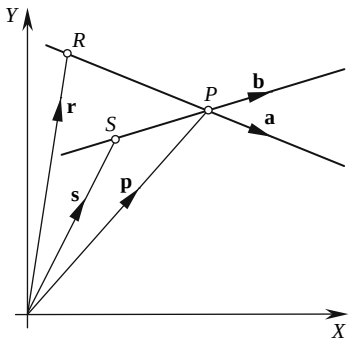


Figure 6.1.

We begin by defining $\mathbf{p}$ as follows for the two lines:

$$\mathbf{p} = \mathbf{r} + \lambda\mathbf{a}$$

$$\mathbf{p} = \mathbf{s} + \varepsilon\mathbf{b}$$

where $P(x_P, y_P)$ is the point of intersection and $\mathbf{p}$ its position vector.
Therefore,

$$\mathbf{r} + \lambda\mathbf{a} = \mathbf{s} + \varepsilon\mathbf{b}. \tag{6.1}$$

To eliminate ε, we multiply Eq. (6.1) throughout by $\mathbf{b}^{\perp}$, as this creates the term $\varepsilon\mathbf{b}^{\perp} \cdot \mathbf{b}$, which is zero:

$$\mathbf{b}^{\perp} \cdot \mathbf{r} + \lambda\mathbf{b}^{\perp} \cdot \mathbf{a} = \mathbf{b}^{\perp} \cdot \mathbf{s} + \varepsilon\mathbf{b}^{\perp} \cdot \mathbf{b} = \mathbf{b}^{\perp} \cdot \mathbf{s}$$

Therefore,

$$\lambda = \frac{\mathbf{b}^{\perp} \cdot (\mathbf{s} - \mathbf{r})}{\mathbf{b}^{\perp} \cdot \mathbf{a}}$$

where

$$\mathbf{b}^{\perp} = -y_b\mathbf{i} + x_b\mathbf{j}.$$

Therefore,

$$\lambda = \frac{(-y_b\mathbf{i} + x_b\mathbf{j}) \cdot ((x_S - x_R)\,\mathbf{i} + (y_S - y_R)\,\mathbf{j})}{(-y_b\mathbf{i} + x_b\mathbf{j}) \cdot (x_a\mathbf{i} + y_a\mathbf{j})}$$

$$\lambda = \frac{x_b\,(y_S - y_R) - y_b\,(x_S - x_R)}{x_b y_a - x_a y_b}. \tag{6.2}$$

To eliminate λ from Eq. (6.1), we multiply throughout by $\mathbf{a}^{\perp}$, as this creates the term $\varepsilon\mathbf{a}^{\perp} \cdot \mathbf{a}$, which is also zero:

$$\mathbf{a}^{\perp} \cdot \mathbf{r} + \lambda\mathbf{a}^{\perp} \cdot \mathbf{a} = \mathbf{a}^{\perp} \cdot \mathbf{s} + \varepsilon\mathbf{a}^{\perp} \cdot \mathbf{b} = \mathbf{a}^{\perp} \cdot \mathbf{r}$$

$$\varepsilon = \frac{\mathbf{a}^{\perp} \cdot (\mathbf{r} - \mathbf{s})}{\mathbf{a}^{\perp} \cdot \mathbf{b}}$$

where

$$\mathbf{a}^{\perp} = -y_a\mathbf{i} + x_a\mathbf{j}.$$

Therefore,

$$\varepsilon = \frac{(-y_a\mathbf{i} + x_a\mathbf{j}) \cdot ((x_R - x_S)\,\mathbf{i} + (y_R - y_S)\,\mathbf{j})}{(-y_a\mathbf{i} + x_a\mathbf{j}) \cdot (x_b\mathbf{i} + y_b\mathbf{j})}$$

$$\varepsilon = \frac{x_a\,(y_S - y_R) - y_a\,(x_S - x_R)}{x_b y_a - x_a y_b}. \tag{6.3}$$

The coordinates of P are

$$x_P = x_R + \lambda x_a \qquad y_P = y_R + \lambda y_a.$$

The lines are parallel if

$$\mathbf{a} = \kappa \mathbf{b} \tag{6.4}$$

where κ is a scalar, and are identical when $\kappa = 1$.
We write Eq. (6.4) as follows:

$$x_a = \kappa x_b$$
$$y_a = \kappa y_b.$$

Cross multiplying gives

$$\kappa x_a y_b = \kappa x_b y_a$$

or

$$x_b y_a - x_a y_b = 0. \tag{6.5}$$

Note that Eq. (6.5) is the denominator in Eqs. (6.2) and (6.3), and if it equals zero, it means that the lines do not intersect.

Let's test the above equations with the example shown in Fig. 6.2.

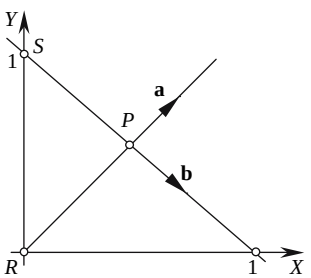


FIGURE 6.2.

The line equations are

$$\mathbf{p} = \mathbf{r} + \lambda \mathbf{a}$$
$$\mathbf{p} = \mathbf{s} + \varepsilon \mathbf{b}$$

where

$$R = (0, 0) \quad \mathbf{a} = \mathbf{i} + \mathbf{j} \quad S = (0, 1) \quad \mathbf{b} = \mathbf{i} - \mathbf{j}.$$

Therefore,

$$\lambda = \frac{(1-0)+(0-0)}{1+1} = \frac{1}{2}$$

and

$$x_P = \frac{1}{2} \qquad y_P = \frac{1}{2}$$

which is correct.

6.3 A line intersecting a circle in $\mathbb{R}^2$

Now let's consider the case of a line intersecting a circle in $\mathbb{R}^2$, for which there are three possible scenarios: the line intersects the circle at two points; the line is tangential to the circle at a single point; or the line does not intersect the circle at all.

The equation of a circle whose center is located at the origin is given by

$$x^2 + y^2 = r^2 \tag{6.6}$$

where r is the radius. But in reality, the circle could be located away from the origin at $(x_C,\ y_C)$, in which case the circle's equation becomes

$$(x - x_C)^2 + (\ y - y_C)^2 = r^2.$$

However, as mentioned above, we can preserve the simplicity of the original circle equation by translating the line equation in the opposite direction. So, for the moment, let's investigate the intersection of a line with Eq. (6.6), assuming that the line equation will be subject to the inverse of the transform applied to the circle.

We begin with the usual line equation

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}$$

where λ is a scalar and $\| \mathbf{v} \| = 1$.

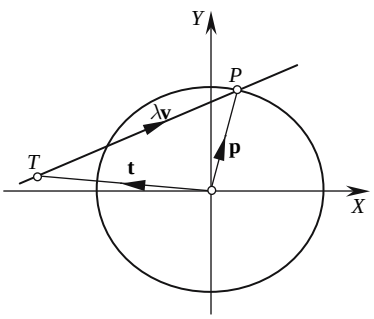


FIGURE 6.3.

Using Eq. (6.6), we can state that

$$x_P^2 + y_P^2 = r^2$$

which is identical to

$$\mathbf{p} \cdot \mathbf{p} = r^2$$

or

$$(\mathbf{t} + \lambda\mathbf{v}) \cdot (\mathbf{t} + \lambda\mathbf{v}) = r^2.$$

Expanding and simplifying gives

$$\mathbf{t} \cdot \mathbf{t} + \lambda 2\mathbf{t} \cdot \mathbf{v} + \lambda^2 \mathbf{v} \cdot \mathbf{v} = r^2.$$

But as $\mathbf{v}$ is a unit vector $\mathbf{v} \cdot \mathbf{v} = 1$, and so, therefore,

$$\lambda^2 + \lambda 2\mathbf{t} \cdot \mathbf{v} + \| \mathbf{t} \|^2 - r^2 = 0$$

which is a quadratic in λ, and solved using

$$\lambda = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = 1 \quad B = 2\mathbf{t} \cdot \mathbf{v} \quad C = \| \mathbf{t} \|^2 - r^2.$$

Therefore,

$$\lambda = \frac{-2\mathbf{t} \cdot \mathbf{v} \pm \sqrt{4(\mathbf{t} \cdot \mathbf{v})^2 - 4\left(\| \mathbf{t} \|^2 - r^2\right)}}{2}$$

$$\lambda = -\mathbf{t} \cdot \mathbf{v} \pm \sqrt{(\mathbf{t} \cdot \mathbf{v})^2 - \| \mathbf{t} \|^2 + r^2}.$$

The discriminant is simplified as follows:

$$\begin{aligned}
(\mathbf{t} \cdot \mathbf{v})^2 - \| \mathbf{t} \|^2 + r^2 &= x_t^2 x_v^2 + y_t^2 y_v^2 + 2x_t x_v y_t y_v - x_t^2 - y_t^2 + r^2 \\
&= x_t^2 \left(x_v^2 - 1\right) + y_t^2 \left(y_v^2 - 1\right) + 2x_t x_v y_t y_v + r^2 \\
&= -x_t^2 y_v^2 - y_t^2 x_v^2 + 2x_t x_v y_t y_v + r^2 \\
&= r^2 - \left(x_t^2 y_v^2 + y_t^2 x_v^2 - 2x_t x_v y_t y_v\right) \\
&= r^2 - |x_v y_t - x_t y_v|^2 \\
(\mathbf{t} \cdot \mathbf{v})^2 - \| \mathbf{t} \|^2 + r^2 &= r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}
\end{aligned}$$

and

$$\lambda = -(x_t x_v + y_t y_v) \pm \sqrt{r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2} \tag{6.7}$$

The value of the discriminant of Eq. (6.7) determines whether the line intersects, is tangential, or misses the circle:

Miss condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 < 0.$$

Tangential condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 = 0.$$

Intersect condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 > 0.$$

Let's test all three conditions using the three lines and circle shown in Fig. 6.4.

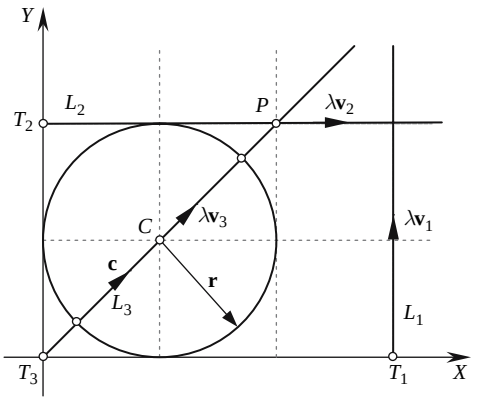


FIGURE 6.4.

Figure 6.4 shows a circle with radius $r = 1$ centered at $C(1, 1)$ and three lines: L_1, L_2, and L_3 that miss, touch, and intersect the circle, respectively.

The lines are

$$\mathbf{p}_1 = \mathbf{t}_1 + \lambda\mathbf{v}_1 \qquad \mathbf{p}_2 = \mathbf{t}_2 + \lambda\mathbf{v}_2 \qquad \mathbf{p}_3 = \mathbf{t}_3 + \lambda\mathbf{v}_3$$

where

$$\begin{aligned} \mathbf{t}_1 &= 3\mathbf{i} & \mathbf{v}_1 &= \mathbf{j} \\ \mathbf{t}_2 &= 2\mathbf{j} & \mathbf{v}_2 &= \mathbf{i} \\ \mathbf{t}_3 &= 0 & \mathbf{v}_3 &= \frac{1}{\sqrt{2}}\mathbf{i} + \frac{1}{\sqrt{2}}\mathbf{j} \end{aligned}$$

and

$$\mathbf{c} = \mathbf{i} + \mathbf{j}.$$

First, we transform the line by the inverse transformation translating the circle away from the origin. In this case:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Therefore,

$$\mathbf{t}' = \mathbf{T}^{-1} \cdot \mathbf{t}$$

$$\begin{bmatrix} x_{t'} \\ y_{t'} \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix}$$

$$x_{t'} = x_t - 1$$
$$y_{t'} = y_t - 1.$$

Another way of stating this using vectors is

$$\mathbf{t}' = \mathbf{t} - \mathbf{c}.$$

Let's substitute the three lines into the following equations:

$$x_P = x_{t'} + \lambda x_v$$
$$y_P = y_{t'} + \lambda y_v$$

where

$$\lambda = -\mathbf{t}' \cdot \mathbf{v} \pm \sqrt{r^2 - \begin{vmatrix} x_v & y_v \\ x_{t'} & y_{t'} \end{vmatrix}^2}$$

and

$$\mathbf{t}' = \mathbf{t} - \mathbf{c}.$$

L_1:

$$\mathbf{t}' = 3\mathbf{i} - (\mathbf{i} + \mathbf{j}) = 2\mathbf{i} - \mathbf{j}$$

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_{t'} & y_{t'} \end{vmatrix}^2 = 1 - \begin{vmatrix} 0 & 1 \\ 2 & -1 \end{vmatrix}^2 = -3.$$

The negative discriminant confirms a miss condition.

L_2:

$$\mathbf{t}' = 2\mathbf{j} - (\mathbf{i}+\mathbf{j}) = -\mathbf{i}+\mathbf{j}$$

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_{t'} & y_{t'} \end{vmatrix}^2 = 1 - \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix}^2 = 0.$$

The zero discriminant confirms a touch condition. Therefore,

$$\lambda = 1$$

and the touch point is

$$x_P = 1 \text{ and } y_P = 2$$

which is correct.

L_3:

$$\mathbf{t}' = 0 - (\mathbf{i}+\mathbf{j}) = -\mathbf{i}-\mathbf{j}$$

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_{t'} & y_{t'} \end{vmatrix}^2 = 1 - \begin{vmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -1 & -1 \end{vmatrix}^2 = 1.$$

The positive discriminant confirms an intersect condition. Therefore,

$$\lambda = \frac{2}{\sqrt{2}} \pm 1 \quad = \sqrt{2}+1 \quad \text{and} \quad \sqrt{2}-1.$$

If we use

$$x_P = x_{t_3} + \lambda x_{v_3} \text{ and } y_P = y_{t_3} + \lambda y_{v_3}$$

the intersection points are

$$\lambda = \sqrt{2}+1 \qquad x_P = \left(\sqrt{2}+1\right)\frac{1}{\sqrt{2}} = 1 + \frac{\sqrt{2}}{2}$$

$$y_P = \left(\sqrt{2}+1\right)\frac{1}{\sqrt{2}} = 1 + \frac{\sqrt{2}}{2}.$$

$$\lambda = \sqrt{2}-1 \qquad x_P = \left(\sqrt{2}-1\right)\frac{1}{\sqrt{2}} = 1 - \frac{\sqrt{2}}{2}$$

$$y_P = \left(\sqrt{2}-1\right)\frac{1}{\sqrt{2}} = 1 - \frac{\sqrt{2}}{2}.$$

The intersection points are $\left(1+\frac{\sqrt{2}}{2},\ 1+\frac{\sqrt{2}}{2}\right)$ and $\left(1-\frac{\sqrt{2}}{2},\ 1-\frac{\sqrt{2}}{2}\right)$, which are correct.

6.4 A line intersecting an ellipse in $\mathbb{R}^2$

section we develop the previous idea of the line-circle intersection with a line-ellipse intersection. We begin with the equation of an ellipse centered at the origin

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \tag{6.8}$$

where a and b are the maximum x- and y-radii, respectively.

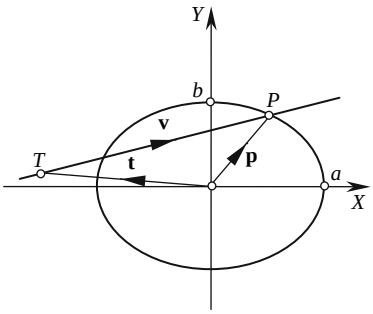


FIGURE 6.5.

Figure 6.5 shows the transformed line and the untransformed ellipse, and we see that the coordinates of P must satisfy Eq. (6.8). Therefore,

$$\frac{(x_t + \lambda x_v)^2}{a^2} + \frac{(y_t + \lambda y_v)^2}{b^2} = 1.$$

Expanding and simplifying gives

$$b^2\left(x_t^2 + \lambda 2 x_t x_v + \lambda^2 x_v^2\right) + a^2\left(y_t^2 + \lambda 2 y_t y_v + \lambda^2 y_v^2\right) = a^2 b^2$$
$$\lambda^2\left(b^2 x_v^2 + a^2 y_v^2\right) + \lambda\left(2b^2 x_t x_v + 2a^2 y_t y_v\right) + a^2 y_t^2 + b^2 x_t^2 - a^2 b^2 = 0$$

which is a quadratic in λ, and solved using

$$\lambda = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = b^2 x_v^2 + a^2 y_v^2$$
$$B = 2b^2 x_t x_v + 2a^2 y_t y_v$$
$$C = a^2 y_t^2 + b^2 x_t^2 - a^2 b^2$$

which reveals the following values for λ:

$$\lambda = \frac{-\left(\frac{x_t x_v}{a^2} + \frac{y_t y_v}{b^2}\right) \pm \sqrt{\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2 b^2}}}{\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2}}. \tag{6.9}$$

The value of the discriminant of Eq. (6.9) determines whether the line intersects, is tangential, or misses the ellipse:

Miss condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2 b^2} < 0.$$

Tangential condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2 b^2} = 0.$$

Intersect condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2 b^2} > 0.$$

We can check the validity of Eq. (6.9) by setting $a = b = r$, which reveals

$$\lambda = \frac{-\left(\frac{x_t x_v}{r^2} + \frac{y_t y_v}{r^2}\right) \pm \sqrt{\frac{x_v^2}{r^2} + \frac{y_v^2}{r^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{r^4}}}{\frac{x_v^2}{r^2} + \frac{y_v^2}{r^2}}$$

$$\lambda = -\left(x_t x_v + y_t y_v\right) \pm \sqrt{r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}$$

which is the same as Eq. (6.7).

6.5 The shortest distance between two skew lines in $\mathbb{R}^3$

Having seen how to cope with two-dimensional lines, circles and ellipses, now let's explore the relationship between two lines in R^3.

We already know that lines in R^2 either intersect or are parallel with one another. However, in R^3 a third option is possible — one where the lines approach one another then recede, allowing a shortest distance to be calculated. Such lines are called *skew* lines, and the shortest distance between them will be on a mutual perpendicular to both lines.

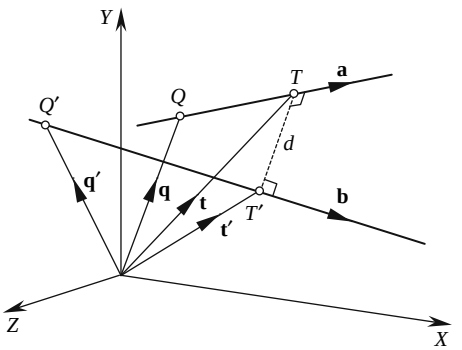


FIGURE 6.6.

Figure 6.6 shows two lines with direction vectors **a** and **b**. The shortest distance d between the lines is the magnitude of the vector $\overrightarrow{TT'}$, which is perpendicular to both lines. Therefore,

$$\overrightarrow{OT} = \mathbf{q} + \lambda\mathbf{a} \tag{6.10}$$

and

$$\overrightarrow{OT'} = \mathbf{q}' + \varepsilon\mathbf{b}. \tag{6.11}$$

But $\overrightarrow{TT'}$ is perpendicular to **a** and **b** and parallel to $\mathbf{a} \times \mathbf{b}$. Therefore,

$$\overrightarrow{TT'} = \frac{d \cdot (\mathbf{a} \times \mathbf{b})}{\| \mathbf{a} \times \mathbf{b} \|}.$$

But

$$\overrightarrow{OT'} = \overrightarrow{OT} + \overrightarrow{TT'}$$

and

$$\overrightarrow{OT'} = \overrightarrow{OT} + \frac{d\,(\mathbf{a} \times \mathbf{b})}{\| \mathbf{a} \times \mathbf{b} \|}. \tag{6.12}$$

We cancel the denominator in Eq. (6.12) by multiplying throughout with $\mathbf{a} \times \mathbf{b}$:

$$(\mathbf{a} \times \mathbf{b}) \cdot \overrightarrow{OT'} = (\mathbf{a} \times \mathbf{b}) \cdot \overrightarrow{OT} + (\mathbf{a} \times \mathbf{b}) \cdot \frac{d\,(\mathbf{a} \times \mathbf{b})}{\| \mathbf{a} \times \mathbf{b} \|}$$

$$(\mathbf{a} \times \mathbf{b}) \cdot \overrightarrow{OT'} = (\mathbf{a} \times \mathbf{b}) \cdot \overrightarrow{OT} + d \| \mathbf{a} \times \mathbf{b} \| . \tag{6.13}$$

We now need to substitute Eqs. (6.10) and (6.11) in Eq. (6.13):

$$(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{q}' + \varepsilon \mathbf{b}) = (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{q} + \lambda \mathbf{a}) + d \| \mathbf{a} \times \mathbf{b} \|$$

$$\mathbf{q}' \cdot (\mathbf{a} \times \mathbf{b}) + \varepsilon \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{q} \cdot (\mathbf{a} \times \mathbf{b}) + \lambda \mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) + d \| \mathbf{a} \times \mathbf{b} \| .$$

As $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a} \times \mathbf{b}$ are mutually perpendicular, $\varepsilon \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0$ and $\lambda \mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = 0$.

Therefore,

$$(\mathbf{q}' - \mathbf{q}) \cdot (\mathbf{a} \times \mathbf{b}) = d \| \mathbf{a} \times \mathbf{b} \|$$

and the shortest distance is

$$d = \frac{| (\mathbf{q} - \mathbf{q}') \cdot (\mathbf{a} \times \mathbf{b}) |}{\| \mathbf{a} \times \mathbf{b} \|}.$$

Let's test this equation with a simple example, as shown in Fig. 6.7. By inspection, the shortest line between the two lines is a perpendicular to $\mathbf{a}$ from the origin, which is $\frac{\sqrt{2}}{2}$ in length.

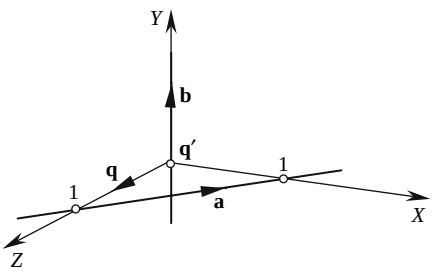


Figure 6.7.

The lines have the following definitions:

$$\mathbf{q} = \mathbf{k} \qquad \mathbf{a} = \mathbf{i} - \mathbf{k} \qquad \mathbf{q}' = 0 \qquad \mathbf{b} = \mathbf{j}.$$

Calculating $\mathbf{a} \times \mathbf{b}$ gives us the following table.

	i	**j**	**k**
a	1	0	–1
b	0	1	0
a × **b**	1	0	1

$$d = \frac{|\,(\mathbf{q}-\mathbf{q}')\cdot(\mathbf{a}\times\mathbf{b})\,|}{\|\,\mathbf{a}\times\mathbf{b}\,\|} = \frac{\mathbf{k}\cdot(\mathbf{i}+\mathbf{k})}{\|\,(\mathbf{i}+\mathbf{k})\,\|} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$

which confirms our prediction.

6.6 Two intersecting lines in $\mathbb{R}^3$

As we have seen in the previous section, two straight lines in $\mathbb{R}^3$ can cross one another but not necessarily intersect. Therefore, in order to calculate a possible intersection, we have to ensure two things:

1. The two lines are not parallel.
2. The two lines touch.

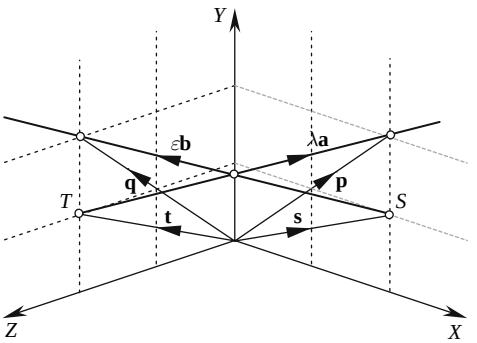


FIGURE 6.8.

We begin by defining the two lines, as shown in Fig. 6.8:

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{a} \quad \text{and} \quad \mathbf{q} = \mathbf{s} + \varepsilon\mathbf{b}$$

where λ and ε are scalars:

$$\mathbf{t} = x_t\mathbf{i} + y_t\mathbf{j} + z_t\mathbf{k} \quad \text{and} \quad \mathbf{s} = x_s\mathbf{i} + y_s\mathbf{j} + z_s\mathbf{k}$$
$$\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k} \quad \text{and} \quad \mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}.$$

Step 1: If $\mathbf{a}\times\mathbf{b} = 0$, the lines are parallel and do not intersect.

Step 2: The distance d between two skew lines is given by

$$d = \frac{\|\,(\mathbf{t}-\mathbf{s})\cdot(\mathbf{a}\times\mathbf{b})\,\|}{\|\,\mathbf{a}\times\mathbf{b}\,\|}.$$

If $(\mathbf{t}-\mathbf{s})\cdot(\mathbf{a}\times\mathbf{b}) \neq 0$, the lines do not intersect.

Step 3: Equate the two line equations:

$$\mathbf{t}+\lambda\mathbf{a}=\mathbf{s}+\varepsilon\mathbf{b},$$
$$(x_t\mathbf{i}+y_t\mathbf{j}+z_t\mathbf{k})+\lambda\,(x_a\mathbf{i}+y_a\mathbf{j}+z_a\mathbf{k})=(x_s\mathbf{i}+y_s\mathbf{j}+z_s\mathbf{k})+\varepsilon\,(x_b\mathbf{i}+y_b\mathbf{j}+z_b\mathbf{k})$$

and collect like components:

$$(x_t-x_s+\lambda x_a-\varepsilon x_b)\,\mathbf{i}+(y_t-y_s+\lambda y_a-\varepsilon y_b)\,\mathbf{j}+(z_t-z_s+\lambda z_a-\varepsilon z_b)\,\mathbf{k}=0.$$

For this vector to be null, its components must vanish. Therefore, we have

$$\lambda x_a-\varepsilon x_b=x_s-x_t$$
$$\lambda y_a-\varepsilon y_b=y_s-y_t$$
$$\lambda z_a-\varepsilon z_b=z_s-z_t$$

which provide values for λ and ε, which, when substituted in the original line equations, reveal the intersection point. For example, the lines in Fig. 6.8 have equations:

$$\mathbf{p}=\mathbf{t}+\lambda\mathbf{a}\quad\text{and}\quad\mathbf{q}=\mathbf{s}+\varepsilon\mathbf{b}\tag{6.14}$$

where

$$\mathbf{t}=\mathbf{j}+2\mathbf{k}\quad\text{and}\quad\mathbf{a}=2\mathbf{i}+\mathbf{j}-2\mathbf{k}$$

and

$$\mathbf{s}=2\mathbf{i}+\mathbf{j}\quad\text{and}\quad\mathbf{b}=-2\mathbf{i}+\mathbf{j}+2\mathbf{k}.$$

The first step is to discover whether the lines are parallel, i.e., if $\mathbf{a}\times\mathbf{b}=0$.

We calculate $\mathbf{a}\times\mathbf{b}$:

	i	**j**	**k**
a	2	1	–2
b	–2	1	2
a × **b**	4	0	4

But as $\mathbf{a}\times\mathbf{b}=4\mathbf{i}+4\mathbf{k}$, the lines are not parallel.

The second step is to discover if the distance d between two skew lines is zero:

$$d=\frac{\|\,(\mathbf{t}-\mathbf{s})\cdot(\mathbf{a}\times\mathbf{b})\,\|}{\|\,\mathbf{a}\times\mathbf{b}\,\|}=\frac{\|\,(-2\mathbf{i}+2\mathbf{k})\cdot(4\mathbf{i}+4\mathbf{k})\,\|}{\|\,4\mathbf{i}+4\mathbf{k}\,\|}$$
$$d=\frac{\|\,(-2\mathbf{i}+2\mathbf{k})\cdot(4\mathbf{i}+4\mathbf{k})\,\|}{\|\,4\mathbf{i}+4\mathbf{k}\,\|}=\frac{-8+8}{\|\,\sqrt{32}\,\|}=0$$

which it is; therefore, the lines do intersect.

The third step is to find the point of intersection using

$$\lambda x_a - \varepsilon x_b = x_s - x_t$$
$$\lambda y_a - \varepsilon y_b = y_s - y_t$$
$$\lambda z_a - \varepsilon z_b = z_s - z_t.$$

Therefore,

$$2\lambda + 2\varepsilon = 2 - 0$$
$$\lambda - \varepsilon = 1 - 1$$
$$-2\lambda - 2\varepsilon = 0 - 2$$

and

$$\lambda + \varepsilon = 1$$
$$\lambda = \varepsilon$$
$$\lambda + \varepsilon = 1.$$

Therefore,

$$\lambda = \frac{1}{2} \quad \text{and} \quad \varepsilon = \frac{1}{2}.$$

Substituting λ and ε in Eq. (6.14), we get

$$\mathbf{p} = (\mathbf{j} + 2\mathbf{k}) + \tfrac{1}{2}(2\mathbf{i} + \mathbf{j} - 2\mathbf{k}) = \mathbf{i} + \tfrac{3}{2}\mathbf{j} + \mathbf{k}.$$

Double checking gives

$$\mathbf{q} = (2\mathbf{i} + \mathbf{j}) + \tfrac{1}{2}(-2\mathbf{i} + \mathbf{j} + 2\mathbf{k}) = \mathbf{i} + \tfrac{3}{2}\mathbf{j} + \mathbf{k}$$

and the point of intersection is $\left(1,\ 1\frac{1}{2},\ 1\right)$, which is correct.

6.7 A line intersecting a plane

There are two configurations of a line and plane: either they intersect or they are parallel, and we must be able to detect both possibilities. The plane could be defined either using a plane equation or using two vectors. As the former is the most probable format, let's proceed with this.

We define the plane equation using the Cartesian form

$$ax + by + cz = d$$

where the normal vector $\mathbf{n}$ is given by

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

We identify a point P on the plane using a position vector $\mathbf{p}$. And as we have seen in previous chapters, we can state

$$\mathbf{n} \cdot \mathbf{p} = d \tag{6.15}$$

which reminds us that if $\mathbf{n}$ is a unit vector, the projection of $\mathbf{p}$ on $\mathbf{n}$ equals the perpendicular distance d from the origin to the plane.

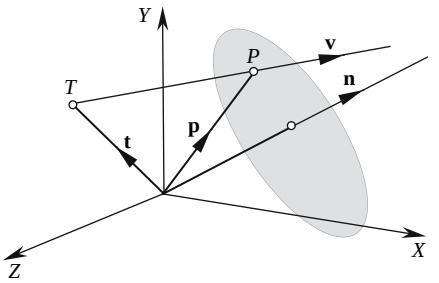


Figure 6.9.

Figure 6.9 shows the scenario described above where the line equation is given by

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}.$$

The line and plane equations could be satisfied for some value of λ, which can be found by substituting $\mathbf{p}$ into Eq. (6.15):

$$\mathbf{n} \cdot (\mathbf{t} + \lambda \mathbf{v}) = d$$
$$\mathbf{n} \cdot \mathbf{t} + \lambda \mathbf{n} \cdot \mathbf{v} = d$$

where

$$\lambda = \frac{d - \mathbf{n} \cdot \mathbf{t}}{\mathbf{n} \cdot \mathbf{v}}.$$

However, if the line is perpendicular to the plane's normal, $\mathbf{n} \cdot \mathbf{v} = 0$ and λ will not have a valid value. Furthermore, if both $\mathbf{n}$ and $\mathbf{v}$ are unit vectors, then

$$\lambda = d - \mathbf{n} \cdot \mathbf{t}.$$

Finally, the position vector for the point of intersection is given by

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}.$$

To test the technique, consider the scenario shown in Fig. 6.10.

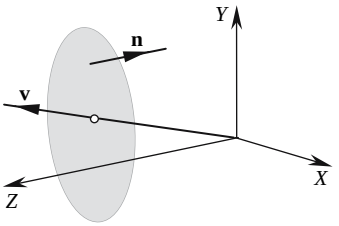


FIGURE 6.10.

The plane equation is given by

$$-\mathbf{k} = -10$$

and the line equation is given by

$$\mathbf{p} = \lambda \mathbf{v}$$

where

$$\mathbf{v} = \mathbf{i} + \mathbf{j} + 10\mathbf{k}.$$

Therefore,

$$\lambda = \frac{d - \mathbf{n} \cdot \mathbf{t}}{\mathbf{n} \cdot \mathbf{v}} = \frac{-10 - (-\mathbf{k}) \cdot (0)}{(-\mathbf{k}) \cdot (\mathbf{i} + \mathbf{j} + 10\mathbf{k})}$$

$$\lambda = \frac{-10}{-10} = 1$$

and the point of intersection is given by

$$\mathbf{p} = \lambda\,(\mathbf{i} + \mathbf{j} + 10\mathbf{k})$$

$$P = (1,\ 1,\ 10)$$

which is correct.

6.8 A line intersecting a sphere

The line-sphere intersection problem is central to ray tracing and is worth investigating. However, before we start exploring the associated geometry, let's pause and remember the case of the line-circle intersection problem in Section 6.3.

Figure 6.3 shows a line intersecting a circle, but Fig. 6.11 is the same diagram with a z-axis and illustrates a line intersecting a sphere. Therefore, surely, the same vector analysis applies, apart from the fact that the vectors are now 3D rather than 2D.

Now the equation of a sphere is

$$x^2 + y^2 + z^2 = r^2 \tag{6.16}$$

where r is the radius and the sphere's center is located at the origin.

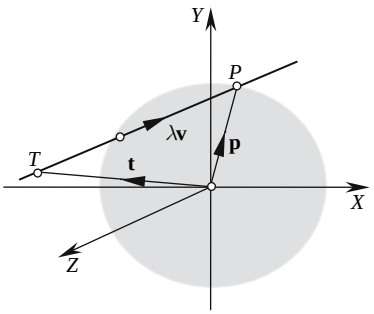


FIGURE 6.11.

We use the same analysis as was used for the line-circle intersection.
The line equation is

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

where λ is a scalar and $\| \mathbf{v} \| = 1$.
Using Eq. (6.16), we can state that

$$x_p^2 + y_p^2 + z_p^2 = r^2$$

which is identical to

$$\mathbf{p} \cdot \mathbf{p} = r^2$$

or

$$(\mathbf{t} + \lambda\mathbf{v}) \cdot (\mathbf{t} + \lambda\mathbf{v}) = r^2.$$

Expanding and simplifying gives

$$\mathbf{t} \cdot \mathbf{t} + \lambda 2\mathbf{t} \cdot \mathbf{v} + \lambda^2 \mathbf{v} \cdot \mathbf{v} = r^2.$$

But as $\mathbf{v}$ is a unit vector, $\mathbf{v} \cdot \mathbf{v} = 1$. Therefore,

$$\lambda^2 + \lambda 2\mathbf{t} \cdot \mathbf{v} + \| \mathbf{t} \|^2 - r^2 = 0$$

which is a quadratic in λ and solved using

$$\lambda = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = 1 \qquad B = 2\mathbf{t} \cdot \mathbf{v} \qquad C = \mathbf{t} \cdot \mathbf{t} - r^2$$

and

$$\lambda = -\mathbf{t}\cdot\mathbf{v} \pm \sqrt{(\mathbf{t}\cdot\mathbf{v})^2 - \|\,\mathbf{t}\,\|^2 + r^2}.$$

Simplifying the discriminant:

$$\begin{aligned}
(\mathbf{t}\cdot\mathbf{v})^2 - \|\,\mathbf{t}\,\|^2 + r^2 &= r^2 + (x_t x_v + y_t y_v + z_t z_v)^2 - \left(x_t^2 + y_t^2 + z_t^2\right) \\
&= r^2 + \begin{pmatrix} x_t^2 x_v^2 + x_t x_v y_t y_v + x_t x_v z_t z_v + y_t^2 y_v^2 + x_t x_v y_t y_v + y_t y_v z_t z_v + \\ z_t^2 z_v^2 + x_t x_v z_t z_v + y_t y_v z_t z_v - x_t^2 - y_t^2 - z_t^2 \end{pmatrix} \\
&= r^2 + \left(x_t^2\left(x_v^2 - 1\right) + y_t^2\left(y_v^2 - 1\right) + z_t^2\left(z_v^2 - 1\right) + 2x_t x_v y_t y_v + 2x_t x_v z_t z_v + 2y_t y_v z_t z_v\right) \\
&= r^2 - \left(x_t^2\left(y_v^2 + z_v^2\right) + y_t^2\left(z_v^2 + x_v^2\right) + z_t^2\left(x_v^2 + y_v^2\right) - 2x_t x_v y_t y_v - 2y_t y_v z_t z_v - 2z_t z_v x_t x_v\right) \\
&= r^2 - \left(x_t^2 y_v^2 + x_t^2 z_v^2 + y_t^2 z_v^2 + y_t^2 x_v^2 + z_t^2 x_v^2 + z_t^2 y_v^2 - 2x_t x_v y_t y_v - 2y_t y_v z_t z_v - 2z_t z_v x_t x_v\right) \\
&= r^2 - \left((x_v y_t - x_t y_v)^2 + (y_v z_t - y_t z_v)^2 + (z_v x_t - z_t x_v)^2\right) \\
&= r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2.
\end{aligned}$$

Therefore,

$$\lambda = -(x_t x_v + y_t y_v + z_t z_v) \pm \sqrt{r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2} \tag{6.17}$$

The value of the discriminant of Eq. (6.17) determines whether the line intersects, is tangential, or misses the sphere:

Miss condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 < 0.$$

Tangential condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 = 0.$$

Intersect condition

$$r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 > 0.$$

Now let's consider a line intersecting an ellipsoid and reveal the similarity in formulas between circle and ellipse with sphere and ellipsoid.

6.9 A line intersecting an ellipsoid

The equation of an ellipsoid is

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1 \tag{6.18}$$

where a, b, and c are the radii on the x-, y-, and z-axes, and the ellipsoid's center is located at the origin, as shown in Fig. 6.12.

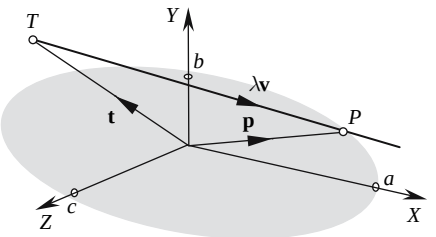


FIGURE 6.12.

The analysis is identical to that used for an ellipse in Section 6.4. So, to develop a general solution for a line-ellipsoid intersection, we subject the ellipsoid to two transforms such as a rotation **R** and a translation **T**. But we know that it is simpler to subject the intersecting line L to the ellipsoid's inverse transforms:

$$L' = \mathbf{R}^{-1}\cdot\mathbf{T}^{-1}\cdot L.$$

So, let's leave the ellipsoid at the origin and use an appropriately transformed line.
Figure 6.12 shows the transformed line and the untransformed ellipsoid, and from this figure we see that the coordinates of P must satisfy Eq. (6.18). Therefore,

$$\frac{(x_t+\lambda x_v)^2}{a^2}+\frac{(y_t+\lambda y_v)^2}{b^2}+\frac{(z_t+\lambda z_v)^2}{c^2}=1.$$

Expanding and simplifying gives

$$\begin{aligned}&b^2c^2(x_t+\lambda x_v)^2+a^2c^2(y_t+\lambda y_v)^2+a^2b^2(z_t+\lambda z_v)^2=a^2b^2c^2\\&\lambda^2(b^2c^2x_v^2+a^2c^2y_v^2+a^2b^2z_v^2)+\lambda(2b^2c^2x_tx_v+2a^2c^2y_ty_v+2a^2b^2z_tz_v)\\&+b^2c^2x_t^2+a^2c^2y_t^2+a^2b^2z_t^2-a^2b^2c^2=0\end{aligned}$$

which is a quadratic in λ and solved using

$$\lambda=\frac{-B\pm\sqrt{B^2-4AC}}{2A}$$

where

$$A = b^2c^2x_v^2 + a^2c^2y_v^2 + a^2b^2z_v^2$$
$$B = 2b^2c^2x_tx_v + 2a^2c^2y_ty_v + 2a^2b^2z_tz_v$$
$$C = b^2c^2x_t^2 + a^2c^2y_t^2 + a^2b^2z_t^2 - a^2b^2c^2$$

and reduces to

$$\lambda = \frac{-\left(\frac{x_tx_v}{a^2} + \frac{y_ty_v}{b^2} + \frac{z_tz_v}{c^2}\right) \pm \sqrt{\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} + \frac{z_v^2}{c^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2b^2} - \frac{\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2}{b^2c^2} - \frac{\begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2}{c^2a^2}}}{\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} + \frac{z_v^2}{c^2}} \tag{6.19}$$

The value of the discriminant of Eq. (6.19) determines whether the line intersects, is tangential, or misses the ellipsoid:

Miss condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} + \frac{z_v^2}{c^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2b^2} - \frac{\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2}{b^2c^2} - \frac{\begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2}{c^2a^2} < 0.$$

Tangential condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} + \frac{z_v^2}{c^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2b^2} - \frac{\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2}{b^2c^2} - \frac{\begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2}{c^2a^2} = 0.$$

Intersect condition

$$\frac{x_v^2}{a^2} + \frac{y_v^2}{b^2} + \frac{z_v^2}{c^2} - \frac{\begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}{a^2b^2} - \frac{\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2}{b^2c^2} - \frac{\begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2}{c^2a^2} > 0.$$

If we relax Eq. (6.19) with $a = b = c = r$, we obtain

$$\lambda = -\mathbf{t}\cdot\mathbf{v} \pm \sqrt{r^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 - \begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 - \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2}$$

which is the equation for a line-sphere intersection.

Before proceeding with the line-cylinder intersection, let's test the above equations for a line-ellipsoid intersection. Eric Haines evaluates a line-ellipsoid intersection in *An Introduction to Ray Tracing* [Glassner, 1989], using a totally different technique. The initial conditions are as follows:

$$\mathbf{t} = 4\mathbf{i} + 5\mathbf{j} - 3\mathbf{k}$$
$$\mathbf{v} = 0.577\mathbf{i} + 0.577\mathbf{j} - 0.577\mathbf{k}$$
$$a = 12 \quad b = 24 \quad c = 8$$

and the ellipsoid's center is

$$(6, 9, -2)\,.$$

The values of λ are –10.3 and 11.1, where the latter identifies the real point of intersection, which turns out to be (10.4, 11.4, −9.4).

Using the technique described above, we begin by subjecting the ray to the inverse of the translation applied to the ellipsoid:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 6 \\ 0 & 1 & 0 & 9 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & 0 & -6 \\ 0 & 1 & 0 & -9 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The untransformed ray is

$$\mathbf{p} = 4\mathbf{i} + 5\mathbf{j} - 3\mathbf{k} + \lambda\,(0.577\mathbf{i} + 0.577\mathbf{j} - 0.577\mathbf{k}) \tag{6.20}$$

and the transformed ray is

$$\mathbf{p}' = -2\mathbf{i} - 4\mathbf{j} - \mathbf{k} + \lambda\,(0.577\mathbf{i} + 0.577\mathbf{j} - 0.577\mathbf{k})\,.$$

Substituting these values in Eq.(6.19), we obtain

$$\lambda = \frac{-\left(\dfrac{-2 \times 0.577}{12^2} - \dfrac{4 \times 0.577}{24^2} - \dfrac{1 \times (-0.577)}{8^2}\right)}{\dfrac{0.577^2}{12^2} + \dfrac{0.577^2}{24^2} + \dfrac{(-0.577)^2}{8^2}}$$

$$\pm \frac{\sqrt{\dfrac{0.577^2}{12^2} + \dfrac{0.577^2}{24^2} + \dfrac{(-0.577)^2}{8^2} - \dfrac{\begin{vmatrix} 0.577 & 0.577 \\ -2 & -4 \end{vmatrix}^2}{12^2 \times 24^2} - \dfrac{\begin{vmatrix} 0.577 & -0.577 \\ -4 & -1 \end{vmatrix}^2}{24^2 \times 8^2} - \dfrac{\begin{vmatrix} -0.577 & 0.577 \\ -1 & -2 \end{vmatrix}^2}{8^2 \times 12^2}}}{\dfrac{0.577^2}{12^2} + \dfrac{0.577^2}{24^2} + \dfrac{(-0.577)^2}{8^2}}$$

$$\lambda = \frac{0.0030052 \pm \sqrt{0.008092024 - 0.000016055 - 0.000225781 - 0.000325125}}{0.008092024}$$

$$\lambda = \frac{0.0030052 \pm \sqrt{0.007525063}}{0.008092024} = -10.3, \quad 11.1.$$

We now substitute the positive value of λ in Eq. (6.20):

$$\mathbf{p} = 4\mathbf{i} + 5\mathbf{j} - 3\mathbf{k} + 11.1\,(0.577\mathbf{i} + 0.577\mathbf{j} - 0.577\mathbf{k})$$

which makes the intersection point, (10.4, 11.4, −9.4).

6.10 A line intersecting a cylinder

Another primitive used in ray tracing is the cylinder. So let's consider a cylinder with radius r and infinite length with its rotational axis aligned with the z-axis, as shown in Fig. 6.13.

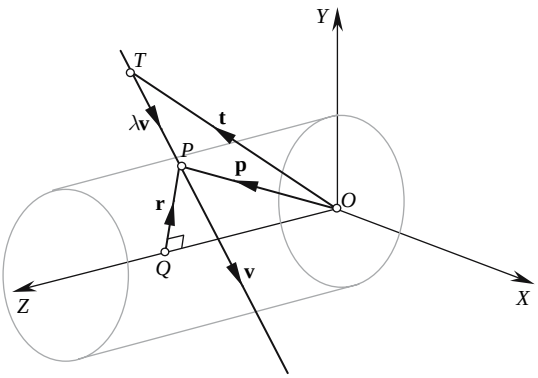


FIGURE 6.13.

The line equation is given by

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v} \tag{6.21}$$

where λ is a scalar and $\| \mathbf{v} \| = 1$.

The point $P\,(x_P, y_P)$ is the point of intersection, and the line QP is constructed such that it is perpendicular to the z-axis, which makes its length equal to r.

We can now state that

$$\| \mathbf{r} \|^2 = x_p^2 + y_p^2 = r^2. \tag{6.22}$$

But from Eq. (6.21), we see that

$$x_p = x_t + \lambda x_v$$
$$y_p = y_t + \lambda y_v$$

and substituting these in Eq. (6.22), we obtain

$$(x_t + \lambda x_v)^2 + (y_t + \lambda y_v)^2 = r^2. \tag{6.23}$$

Expanding Eq. (6.23), we obtain

$$\lambda^2\left(x_v^2+y_v^2\right)+\lambda\left(2x_vx_t+2y_vy_t\right)+x_t^2+y_t^2-r^2=0$$

which is a quadratic in λ and solved using

$$\lambda=\frac{-B\pm\sqrt{B^2-4AC}}{2A}$$

where

$$A=x_v^2+y_v^2 \qquad B=2\left(x_vx_t+y_vy_t\right) \qquad C=x_t^2+y_t^2-r^2$$

and can be represented by

$$\lambda=\frac{-\left(x_vx_t+y_vy_t\right)\pm\sqrt{r^2\left(x_v^2+y_v^2\right)-\begin{vmatrix}x_v & y_v\\ x_t & y_t\end{vmatrix}^2}}{x_v^2+y_v^2}. \tag{6.24}$$

The value of the discriminant of Eq. (6.24) determines whether the line intersects, is tangential, or misses the cylinder:

Miss condition

$$r^2\left(x_v^2+y_v^2\right)-\begin{vmatrix}x_v & y_v\\ x_t & y_t\end{vmatrix}^2<0.$$

Tangential condition

$$r^2\left(x_v^2+y_v^2\right)-\begin{vmatrix}x_v & y_v\\ x_t & y_t\end{vmatrix}^2=0.$$

Intersect condition

$$r^2\left(x_v^2+y_v^2\right)-\begin{vmatrix}x_v & y_v\\ x_t & y_t\end{vmatrix}^2>0.$$

It is highly unlikely that the cylinder will be aligned with the z-axis. And as we saw with the ellipsoid, it is simpler to subject the intersecting line L to the cylinder's inverse transforms:

$$L'=\mathbf{S}^{-1}\cdot\mathbf{R}^{-1}\cdot\mathbf{T}^{-1}\cdot L.$$

So, let's leave the cylinder aligned with the z-axis and use an appropriately transformed line. If the line equation is

$$\mathbf{p}=\mathbf{t}+\lambda\mathbf{v}$$

then

$$\mathbf{v}' = \mathbf{R}^{-1} \cdot \mathbf{S}^{-1} \cdot \mathbf{v}$$
$$\mathbf{t}' = \mathbf{R}^{-1} \cdot \mathbf{T}^{-1} \cdot \mathbf{t}$$

and

$$\mathbf{p} = \mathbf{t}' + \lambda \mathbf{v}'.$$

Note that $\mathbf{v}$ is not subjected to the translation transformation. Therefore, Eq. (6.24) becomes

$$\lambda = \frac{-(x_{v'}x_{t'} + y_{v'}y_{t'}) \pm \sqrt{r^2(x_{v'}^2 + y_{v'}^2) - \begin{vmatrix} x_{v'} & y_{v'} \\ x_{t'} & y_{t'} \end{vmatrix}^2}}{x_{v'}^2 + y_{v'}^2}. \tag{6.25}$$

Let's test Eq. (6.25) with the following conditions:
If the scale transform is

$$\mathbf{S} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ then } \mathbf{S}^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If the rotate transform is

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 90^\circ & -\sin 90^\circ & 0 \\ 0 & \sin 90^\circ & \cos 90^\circ & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ then } \mathbf{R}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 90^\circ & \sin 90^\circ & 0 \\ 0 & -\sin 90^\circ & \cos 90^\circ & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If the translate transform is

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ then } \mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -10 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

If the ray is defined as

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}$$

where

$$\mathbf{t} = 0$$

and

$$\mathbf{v} = \mathbf{j} + \mathbf{k}$$

we have the scenario shown in Fig. 6.14, where the cylinder has been subjected to the transforms rather than the line. From Fig. 6.14 we can predict that the ray intersection points are given by (0, 8, 8) and (0, 12, 12). Let's now confirm these predictions by transforming the ray.

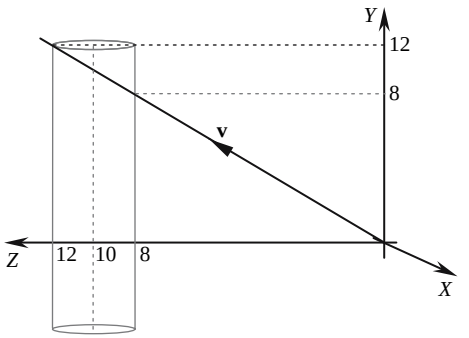


FIGURE 6.14.

Calculate

$$\mathbf{S}^{-1}\cdot\mathbf{R}^{-1}=\begin{bmatrix}\frac{1}{2}&0&0&0\\0&\frac{1}{2}&0&0\\0&0&\frac{1}{2}&0\\0&0&0&1\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0\\0&0&1&0\\0&-1&0&0\\0&0&0&1\end{bmatrix}=\begin{bmatrix}\frac{1}{2}&0&0&0\\0&0&\frac{1}{2}&0\\0&-\frac{1}{2}&0&0\\0&0&0&1\end{bmatrix}$$

and

$$\mathbf{S}^{-1}\cdot\mathbf{R}^{-1}\cdot\mathbf{T}^{-1}=\begin{bmatrix}\frac{1}{2}&0&0&0\\0&0&\frac{1}{2}&0\\0&-\frac{1}{2}&0&0\\0&0&0&1\end{bmatrix}\cdot\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&-10\\0&0&0&1\end{bmatrix}=\begin{bmatrix}\frac{1}{2}&0&0&0\\0&0&\frac{1}{2}&-5\\0&-\frac{1}{2}&0&0\\0&0&0&1\end{bmatrix}.$$

If the ray is given by

$$\mathbf{p}=\mathbf{t}+\lambda\mathbf{v} \tag{6.26}$$

where

$$\mathbf{t}=0 \text{ and } \mathbf{v}=\mathbf{j}+\mathbf{k}$$

then

$$\mathbf{t}'=\begin{bmatrix}\frac{1}{2}&0&0&0\\0&0&\frac{1}{2}&-5\\0&-\frac{1}{2}&0&0\\0&0&0&1\end{bmatrix}\cdot\begin{bmatrix}0\\0\\0\\1\end{bmatrix}=\begin{bmatrix}0\\-5\\0\\1\end{bmatrix}$$

and

$$\mathbf{v}' = \begin{bmatrix} \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}.$$

Substituting $\mathbf{t}'$ and $\mathbf{v}'$ in Eq. (6.25), we get

$$\lambda = \frac{2\frac{1}{2} \pm \sqrt{\frac{1}{4} - \begin{vmatrix} 0 & \frac{1}{2} \\ 0 & -5 \end{vmatrix}^2}}{\frac{1}{4}} = \frac{2\frac{1}{2} \pm \frac{1}{2}}{\frac{1}{4}} = 8 \text{ and } 12.$$

Substituting λ in Eq. (6.26), we get

$$\mathbf{p} = 0 + 8(\mathbf{j} + \mathbf{k}) \text{ and } \mathbf{p} = 0 + 12(\mathbf{j} + \mathbf{k})$$

which makes the intersection points (0, 8, 8) and (0, 12, 12), as predicted!

What has been covered so far assumes that the cylinder has an infinite extent along the z-axis, which is not very common. Normally, cylinders have a finite length, and a mechanism is required to set these physical limits. One convenient method is to identify the minimum and maximum z-extents of the cylinder, $z_{\min}$ and $z_{\max}$, respectively. And when we calculate the two possible intersection values for λ, e.g., λ_1 and λ_2, we can test whether the z component of $\mathbf{p}$ is within the defined range:

$$z_1 = z_t + \lambda_1 z_v$$
$$z_2 = z_t + \lambda_2 z_v$$

by making sure that

$$z_{\min} < (z_1, z_2) < z_{\max}.$$

Another assumption with the above analysis is that the cylinder is open at either end. If end caps are required, then plane equations can be placed at $z_{\min}$ and $z_{\max}$ as shown in Fig. 6.15.

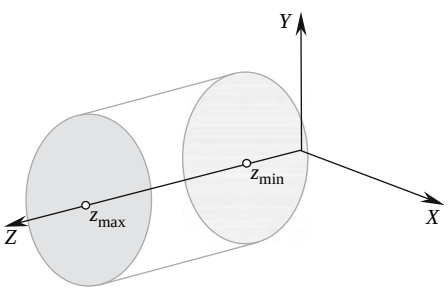


FIGURE 6.15.

We also need to recognise the conditions when such an intersection can arise. An intersection with the $z_{\min}$ end cap occurs when z_1 and z_2 are either side of the $z_{\min}$ plane. Similarly, an intersection with the $z_{\max}$ end cap occurs when z_1 and z_2 are either side of the $z_{\max}$ plane.

We have already examined the intersection of a line with a plane in Section 6.7 and discovered that given a line

$$\mathbf{p} = \mathbf{t} + \lambda \mathbf{v}$$

and a plane equation

$$ax + by + cz = d$$

where

$$\mathbf{n} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$
$$\lambda = \frac{-(\mathbf{n} \cdot \mathbf{t} + d)}{\mathbf{n} \cdot \mathbf{v}}.$$

If we place a plane at $z_{\min}$, we have

$$d = z_{\min} \text{ and } \mathbf{n} = -\mathbf{k}$$

and

$$\lambda_3 = \frac{-(-z_T + z_{\min})}{-z_v} = \frac{z_{\min} - z_T}{z_v}.$$

λ_3 can now be used to reveal the intersection point using

$$\mathbf{p} = \mathbf{t} + \lambda_3 \mathbf{v}.$$

If we place a plane at $z_{\max}$, we have

$$d = z_{\max} \text{ and } \mathbf{n} = -\mathbf{k}$$

and

$$\lambda_4 = \frac{-(-z_T + z_{\max})}{-z_v} = \frac{z_{\max} - z_T}{z_v}.$$

λ_4 can now be used to reveal the intersection point using

$$\mathbf{p} = \mathbf{t} + \lambda_4 \mathbf{v}.$$

Note that both end cap normals are pointing towards the origin.

6.11 A line intersecting a cone

Another ray-tracing primitive is the cone, and once again this is normally aligned, like the cylinder, with the z-axis, as shown in Fig. 6.16. The simplest form of the cone equation is

$$z^2 = x^2 + y^2$$

which makes the internal angle of the apex 90°.

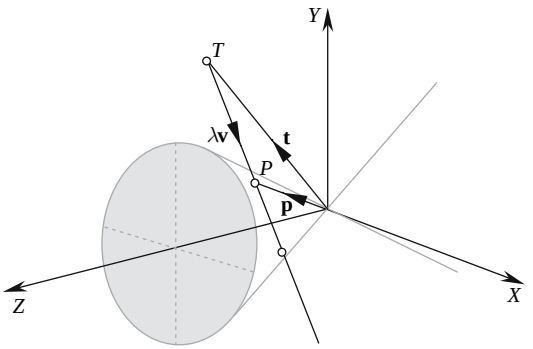


FIGURE 6.16.

If the line equation is given by

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

where λ is a scalar and $\| \mathbf{v} \| = 1$.

Then for an intersection at P, we have

$$z_P^2 = x_P^2 + y_P^2$$

which means that

$$(z_t + \lambda z_v)^2 = (x_t + \lambda x_v)^2 + (y_t + \lambda y_v)^2 .$$

Expanding and simplifying gives

$$z_t^2 + 2\lambda z_t z_v + \lambda^2 z_v^2 = x_t^2 + 2\lambda x_t x_v + \lambda^2 x_v^2 + y_t^2 + 2\lambda y_t y_v + \lambda^2 y_v^2$$
$$0 = \lambda^2 \left(x_v^2 + y_v^2 - z_v^2\right) + \lambda \left(2x_t x_v + 2y_t y_v - 2z_t z_v\right) + x_t^2 + y_t^2 - z_t^2$$

which is a quadratic in λ and solved using

$$\lambda = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

where

$$A = x_v^2 + y_v^2 - z_v^2 \qquad B = 2x_t x_v + 2y_t y_v - 2z_t z_v \qquad C = x_t^2 + y_t^2 - z_t^2.$$

This reduces to

$$\lambda = \frac{-(x_t x_v + y_t y_v - z_t z_v) \pm \sqrt{\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 + \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2}}{x_v^2 + y_v^2 - z_v^2} \tag{6.27}$$

The value of the discriminant of Eq. (6.27) determines whether the line intersects, is tangential, or misses the cone:

Miss condition

$$\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 + \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 < 0.$$

Tangential condition

$$\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 + \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 = 0.$$

Intersect condition

$$\begin{vmatrix} y_v & z_v \\ y_t & z_t \end{vmatrix}^2 + \begin{vmatrix} z_v & x_v \\ z_t & x_t \end{vmatrix}^2 - \begin{vmatrix} x_v & y_v \\ x_t & y_t \end{vmatrix}^2 > 0.$$

If the denominator of Eq. (6.27) is zero, the ray is parallel with the cone's side.

Limits to the cone's length can be set using the same technique for the cylinder, where $z_{\min}$ and $z_{\max}$ define the z-extent of the cone. These can be negative or positive and determine whether the cone is aligned with the negative or positive z-axis, respectively.

If the ray intersects the cone twice, there will be two values of $\lambda : \lambda_1$ and λ_2, which generate the two z coordinates of the intersections:

$$z_1 = z_t + \lambda_1 z_v$$
$$z_2 = z_t + \lambda_2 z_v.$$

Once again, to ensure that the intersections occur in the isolated part of the cone, we test for

$$z_{\min} < (z_1,\ z_2) < z_{\max}.$$

By making either $z_{\min}$ or $z_{\max}$ equal to zero retains the cone's apex. However, if they are both negative or both positive, a truncated cone is created. And if the cone requires end caps, the same technique used for capping cylinders can be applied to cones.

6.12 A line intersecting a triangle

Triangles play an important role in modeling CG objects, and the line-triangle intersection problem is important to the design of renderers. Triangles are planar, and it is easy to compute whether, and where, a line intersects the plane shared by the triangle. What is not so easy is proving that the intersection is within the triangle's boundary. Fortunately for us, in 1827, August Ferdinand Möbius [1790–1868] published *Der barycentrische Calcul*, which formalized the analysis of centers of gravity and also laid the foundations for vectors. Today the work of Möbius is called barycentric coordinates.

Barycentric coordinates provide the key to unlocking this problem, and if you have never used them, refer to the author's book *Mathematics for Computer Graphics* [Vince, 2006].

We begin by defining triangle $\Delta P_1P_2P_3$ with vertices $P_1(x_1, y_1, z_1)$, $P_2(x_2, y_2, z_2)$, and $P_3(x_3, y_3, z_3)$ and their corresponding position vectors $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. This scenario is shown in Fig. 6.17.

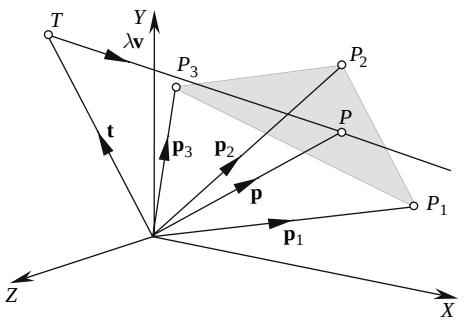


Figure 6.17.

We can state that any point $P(x_P, y_P, z_P)$ on triangle $\Delta P_1P_2P_3$ is given by

$$\mathbf{p}(r, s) = q\mathbf{p}_1 + r\mathbf{p}_2 + s\mathbf{p}_3$$

where $\mathbf{p}$ is P's position vector. The multipliers q, r, and s are barycentric coordinates and satisfy the following conditions:

$$r + s + (1 - r - s) = 1 \qquad (q, r, s) \geq 0$$

which makes

$$q = 1 - r - s.$$

If the intersecting ray is given by

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

where λ is a scalar, then

$$\begin{aligned}\mathbf{t}+\lambda\mathbf{v} &= (1-r-s)\,\mathbf{p}_1 + r\mathbf{p}_2 + s\mathbf{p}_3\\ \mathbf{t}+\lambda\mathbf{v} &= \mathbf{p}_1 - r\mathbf{p}_1 - s\mathbf{p}_1 + r\mathbf{p}_2 + s\mathbf{p}_3\\ \mathbf{t}+\lambda\mathbf{v} &= \mathbf{p}_1 + r\,(\mathbf{p}_2-\mathbf{p}_1) + s\,(\mathbf{p}_3-\mathbf{p}_1)\\ \mathbf{t}-\mathbf{p}_1 &= -\lambda\mathbf{v} + r\,(\mathbf{p}_2-\mathbf{p}_1) + s\,(\mathbf{p}_3-\mathbf{p}_1)\end{aligned}$$

and

$$\mathbf{t}-\mathbf{p}_1 = \begin{bmatrix}-\mathbf{v} & \mathbf{p}_2-\mathbf{p}_1 & \mathbf{p}_3-\mathbf{p}_1\end{bmatrix}\begin{bmatrix}\lambda\\ r\\ s\end{bmatrix}. \tag{6.28}$$

Solving Eq. (6.28) reveals the value of λ, r, and s. The barycentric coordinates r and s tell us whether the point is on the triangle, and λ is the position along $\mathbf{v}$. First, though, we need to solve Eq. (6.28).

As an aside, consider the simultaneous equations

$$\begin{aligned}c_1 &= a_1x + b_1y\\ c_2 &= a_2x + b_2y\end{aligned}$$

which can be solved as follows:

$$\begin{aligned}b_2c_1 &= a_1b_2x + b_1b_2y\\ -b_1c_2 &= -a_2b_1x - b_1b_2y\end{aligned}$$

adding

$$b_2c_1 - b_1c_2 = a_1b_2x - a_2b_1x = (a_1b_2 - a_2b_1)\,x$$

and

$$x = \frac{b_2c_1 - b_1c_2}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix}c_1 & b_1\\ c_2 & b_2\end{vmatrix}}{\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}}.$$

Similarly, we have

$$\begin{aligned}-a_2c_1 &= -a_1a_2x - a_2b_1y\\ a_1c_2 &= a_1a_2x + a_1b_2y\end{aligned}$$

adding

$$a_1c_2 - a_2c_1 = a_1b_2y - a_2b_1y = (a_1b_2 - a_2b_1)\,y$$

$$y = \frac{a_1c_2 - a_2c_1}{a_1b_2 - a_2b_1} = \frac{\begin{vmatrix}a_1 & c_1\\ a_2 & c_2\end{vmatrix}}{\begin{vmatrix}a_1 & b_1\\ a_2 & b_2\end{vmatrix}}.$$

This symmetry was discovered by the Swiss mathematician Gabriel Cramer [1704–1752] and is now known as Cramer's rule. However, it only works when the determinant in the denominator is non-zero.

Similarly, for three simultaneous equations:

$$d_1 = a_1x + b_1y + c_1z$$
$$d_2 = a_2x + b_2y + c_2z$$
$$d_3 = a_3x + b_3y + c_3z$$

we can state

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

and

$$x = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \qquad y = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}} \qquad z = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}}.$$

Before applying Cramer's rule, let's tidy up Eq. (6.28) as follows:

$$\mathbf{k} = \mathbf{t} - \mathbf{p}_1 \qquad \mathbf{p}_{21} = \mathbf{p}_2 - \mathbf{p}_1 \qquad \mathbf{p}_{31} = \mathbf{p}_3 - \mathbf{p}_1$$

Therefore,

$$\mathbf{k} = \begin{bmatrix} -\mathbf{v} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{bmatrix} \begin{bmatrix} \lambda \\ r \\ s \end{bmatrix} \tag{6.29}$$

and

$$\lambda = \frac{\begin{vmatrix} x_k & x_{21} & x_{31} \\ y_k & y_{21} & y_{31} \\ z_k & z_{21} & z_{31} \end{vmatrix}}{DET} \qquad r = \frac{\begin{vmatrix} -x_v & x_k & x_{31} \\ -y_v & y_k & y_{31} \\ -z_v & z_k & z_{31} \end{vmatrix}}{DET} \qquad s = \frac{\begin{vmatrix} -x_v & x_{21} & x_k \\ -y_v & y_{21} & y_k \\ -z_v & z_{21} & z_k \end{vmatrix}}{DET} \tag{6.30}$$

where

$$DET = \begin{vmatrix} -x_v & x_{21} & x_{31} \\ -y_v & y_{21} & y_{31} \\ -z_v & z_{21} & z_{31} \end{vmatrix}.$$

In their book *Real-Time Rendering*, Tomas Akenine-Möller and Eric Haines [Akenine-Möller, 2002] take the above solution one stage further by exploiting the fact that the determinant

$$\begin{vmatrix} \mathbf{a} & \mathbf{b} & \mathbf{c} \end{vmatrix} = -\begin{vmatrix} \mathbf{a} & \mathbf{c} & \mathbf{b} \end{vmatrix} = -(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{b} = -(\mathbf{c} \times \mathbf{b}) \cdot \mathbf{a}.$$

You may recall that we covered this in Section 2.9.1 and proved that the box product

$$\begin{vmatrix} x_a & y_a & z_a \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\cdot(\mathbf{c}\times\mathbf{a}) = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = -\mathbf{a}\cdot(\mathbf{c}\times\mathbf{b}) = -\mathbf{b}\cdot(\mathbf{a}\times\mathbf{c}) = -\mathbf{c}\cdot(\mathbf{b}\times\mathbf{a}).$$

Applying these relationships to Eq. (6.29), we obtain

$$\begin{aligned} \lambda &= \frac{\begin{vmatrix} \mathbf{k} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{vmatrix}}{\begin{vmatrix} -\mathbf{v} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{vmatrix}} = \frac{(\mathbf{k}\times\mathbf{p}_{21})\cdot\mathbf{p}_{31}}{(\mathbf{v}\times\mathbf{p}_{31})\cdot\mathbf{p}_{21}} \\ r &= \frac{\begin{vmatrix} -\mathbf{v} & \mathbf{k} & \mathbf{p}_{31} \end{vmatrix}}{\begin{vmatrix} -\mathbf{v} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{vmatrix}} = \frac{(\mathbf{v}\times\mathbf{p}_{31})\cdot\mathbf{k}}{(\mathbf{v}\times\mathbf{p}_{31})\cdot\mathbf{p}_{21}} \\ s &= \frac{\begin{vmatrix} -\mathbf{v} & \mathbf{p}_{21} & \mathbf{k} \end{vmatrix}}{\begin{vmatrix} -\mathbf{v} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{vmatrix}} = \frac{(\mathbf{k}\times\mathbf{p}_{21})\cdot\mathbf{v}}{(\mathbf{v}\times\mathbf{p}_{31})\cdot\mathbf{p}_{21}}. \end{aligned} \tag{6.31}$$

Note that the above equations for λ, r, and s contain two common terms: $\mathbf{k}\times\mathbf{p}_{21}$ and $\mathbf{v}\times\mathbf{p}_{31}$.

If these are computed separately, we can write

$$\lambda = \frac{\mathbf{m}\cdot\mathbf{p}_{31}}{\mathbf{n}\cdot\mathbf{p}_{21}} \qquad r = \frac{\mathbf{n}\cdot\mathbf{k}}{\mathbf{n}\cdot\mathbf{p}_{21}} \qquad s = \frac{\mathbf{m}\cdot\mathbf{v}}{\mathbf{n}\cdot\mathbf{p}_{21}},$$

where

$$\mathbf{m} = \mathbf{k}\times\mathbf{p}_{21} \qquad \mathbf{n} = \mathbf{v}\times\mathbf{p}_{31}.$$

Let's put the above theory into practice using a triangle formed from the points $P_1\,(0, 2, 0)$, $P_2\,(0, 0, 4)$, and $P_3\,(3, 1, 2)$ with a ray

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

where

$$\mathbf{t} = 4\mathbf{j} \qquad \mathbf{v} = -\mathbf{j} + \mathbf{k}$$

which will intersect the triangle at P_2, as shown in Fig. 6.18.

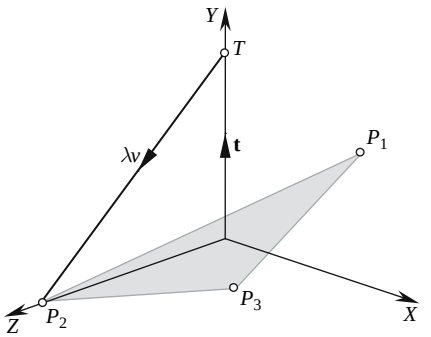


FIGURE 6.18.

Therefore,

$$\mathbf{k} = \mathbf{t} - \mathbf{p}_1 = 4\mathbf{j} - 2\mathbf{j} = 2\mathbf{j}$$
$$\mathbf{p}_{21} = \mathbf{p}_2 - \mathbf{p}_1 = 4\mathbf{k} - 2\mathbf{j} = -2\mathbf{j} + 4\mathbf{k}$$
$$\mathbf{p}_{31} = \mathbf{p}_3 - \mathbf{p}_1 = 3\mathbf{i} + \mathbf{j} + 2\mathbf{k} - 2\mathbf{j} = 3\mathbf{i} - \mathbf{j} + 2\mathbf{k}$$

and using Eq. (6.30), we obtain

$$DET = \begin{vmatrix} 0 & 0 & 3 \\ 1 & -2 & -1 \\ -1 & 4 & 2 \end{vmatrix} = 6$$
$$\lambda = \frac{1}{6}\begin{vmatrix} 0 & 0 & 3 \\ 2 & -2 & -1 \\ 0 & 4 & 2 \end{vmatrix} = 4$$
$$r = \frac{1}{6}\begin{vmatrix} 0 & 0 & 3 \\ 1 & 2 & -1 \\ -1 & 0 & 2 \end{vmatrix} = 1$$
$$s = \frac{1}{6}\begin{vmatrix} 0 & 0 & 0 \\ 1 & -2 & 2 \\ -1 & 4 & 0 \end{vmatrix} = 0$$

$r = 1$ confirms that the ray intersects the triangle at P_2. Similarly, with $\lambda = 4$, the ray intersects the point $\mathbf{p} = 4\mathbf{j} + 4[-\mathbf{j} + \mathbf{k}] = 4\mathbf{k}$, which is P_2.

Using Eq. (6.31), we find that

$$\mathbf{m} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 2 & 0 \\ 0 & -2 & 4 \end{vmatrix} = 8\mathbf{i}$$
$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & -1 & 1 \\ 3 & -1 & 2 \end{vmatrix} = -\mathbf{i} + 3\mathbf{j} + 3\mathbf{k}$$
$$\mathbf{n} \cdot \mathbf{p}_{21} = (-\mathbf{i} + 3\mathbf{j} + 3\mathbf{k}) \cdot (-2\mathbf{j} + 4\mathbf{k}) = 6$$
$$\lambda = \frac{1}{6}(8\mathbf{i}) \cdot (3\mathbf{i} - \mathbf{j} + 2\mathbf{k}) = 4$$
$$\mathbf{r} = \frac{1}{6}(-\mathbf{i} + 3\mathbf{j} + 3\mathbf{k}) \cdot (2\mathbf{j}) = 1$$
$$\mathbf{s} = \frac{1}{6}(8\mathbf{i}) \cdot (-\mathbf{j} + \mathbf{k}) = 0$$

which are identical to the above results.

Now let's try another ray:

$$\mathbf{t} = 4\mathbf{j}, \quad \mathbf{v} = \mathbf{i} - \mathbf{j} + \mathbf{k}.$$

Using Eq. (6.30), we obtain

$$\mathbf{k} = 2\mathbf{j} \quad \text{(unchanged)}$$

$$DET = \begin{vmatrix} -1 & 0 & 3 \\ 1 & -2 & -1 \\ -1 & 4 & 2 \end{vmatrix} = 6$$

$$\lambda = \frac{1}{6}\begin{vmatrix} 0 & 0 & 3 \\ 2 & -2 & -1 \\ 0 & 4 & 2 \end{vmatrix} = 4$$

$$r = \frac{1}{6}\begin{vmatrix} -1 & 0 & 3 \\ 1 & 2 & -1 \\ -1 & 0 & 2 \end{vmatrix} = \frac{1}{3}$$

$$s = \frac{1}{6}\begin{vmatrix} -1 & 0 & 0 \\ 1 & -2 & 2 \\ -1 & 4 & 0 \end{vmatrix} = 1\frac{1}{3}.$$

Because the value of $s > 1$, the intersection point is outside the triangle.

Let's make sure that we obtain the same result using Eq. (6.31):

$$\mathbf{m} = 8\mathbf{i} \quad \text{(unchanged)}$$

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -1 & 1 \\ 3 & -1 & 2 \end{vmatrix} = -\mathbf{i} + \mathbf{j} + 2\mathbf{k}$$

$$\mathbf{n} \cdot \mathbf{p}_{21} = (-\mathbf{i} + \mathbf{j} + 2\mathbf{k}) \cdot (-2\mathbf{j} + 4\mathbf{k}) = 6$$

$$\lambda = \frac{1}{6}(8\mathbf{i}) \cdot (3\mathbf{i} - \mathbf{j} + 2\mathbf{k}) = 4$$

$$\mathbf{r} = \frac{1}{6}(-\mathbf{i} + \mathbf{j} + 2\mathbf{k}) \cdot (2\mathbf{j}) = \frac{1}{3}$$

$$\mathbf{s} = \frac{1}{6}(8\mathbf{i}) \cdot (\mathbf{i} + -\mathbf{j} + \mathbf{k}) = 1\frac{1}{3}$$

which confirm the previous values.

Finally, let's select another ray that hopefully intersects the triangle:

$$\mathbf{t} = 4\mathbf{j} \qquad \mathbf{v} = 2\mathbf{i} - 4\mathbf{j} + 2\mathbf{k}.$$

Using Eq. (6.30), we obtain

$$\mathbf{k} = 2\mathbf{j} \quad \text{(unchanged)}$$

$$DET = \begin{vmatrix} -2 & 0 & 3 \\ 4 & -2 & -1 \\ -2 & 4 & 2 \end{vmatrix} = 36$$

$$\lambda = \frac{1}{36}\begin{vmatrix} 0 & 0 & 3 \\ 2 & -2 & -1 \\ 0 & 4 & 2 \end{vmatrix} = \frac{2}{3}$$

$$r = \frac{1}{36}\begin{vmatrix} -2 & 0 & 3 \\ 4 & 2 & -1 \\ -2 & 0 & 2 \end{vmatrix} = \frac{1}{9}$$

$$s = \frac{1}{36}\begin{vmatrix} -2 & 0 & 0 \\ 4 & -2 & 2 \\ -2 & 4 & 0 \end{vmatrix} = \frac{4}{9}$$

$r = \frac{1}{9}$ and $s = \frac{4}{9}$ confirms that the ray intersects the triangle. The point of intersection is given by

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v} = 4\mathbf{j} + \frac{2}{3}(2\mathbf{i} - 4\mathbf{j} + 2\mathbf{k})$$

$$= 4\mathbf{j} + \frac{4}{3}\mathbf{i} - \frac{8}{3}\mathbf{j} + \frac{4}{3}\mathbf{k}$$

$$P = \left(\frac{4}{3}, \frac{4}{3}, \frac{4}{3}\right).$$

Using Eq. (6.31), we find that

$$\mathbf{k} = 2\mathbf{j} \quad \text{(unchanged)}$$

$$\mathbf{m} = 8\mathbf{i} \quad \text{(unchanged)}$$

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 2 & -4 & 2 \\ 3 & -1 & 2 \end{vmatrix} = -6\mathbf{i} + 2\mathbf{j} + 10\mathbf{k}$$

$$\mathbf{n} \cdot \mathbf{p}_{21} = (-6\mathbf{i} + 2\mathbf{j} + 10\mathbf{k}) \cdot (-2\mathbf{j} + 4\mathbf{k}) = 36$$

$$\lambda = \frac{1}{36}(8\mathbf{i}) \cdot (3\mathbf{i} - \mathbf{j} + 2\mathbf{k}) = \frac{2}{3}$$

$$r = \frac{1}{36}(-6\mathbf{i} + 2\mathbf{j} + 10\mathbf{k}) \cdot (2\mathbf{j}) = \frac{1}{9}$$

$$s = \frac{1}{36}(8\mathbf{i}) \cdot (2\mathbf{i} - 4\mathbf{j} + 2\mathbf{k}) = \frac{4}{9}$$

which are identical to the previous results.

6.13 A point inside a triangle

Having discovered how to test whether a line intersects a triangle, it is an opportune moment to describe how to test whether a point is inside or outside a triangle. Once again, we draw upon barycentric coordinates to reveal the solution.

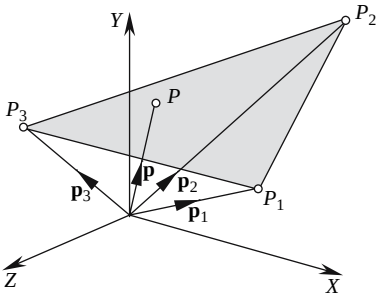


FIGURE 6.19.

Figure 6.19 shows a triangle $\Delta P_1P_2P_3$ with position vectors $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. Given a point P, which resides on the plane containing P_1, P_2, and P_3, we can state that

$$\mathbf{p} = r\mathbf{p}_1 + s\mathbf{p}_2 + t\mathbf{p}_3 \tag{6.32}$$

where $r + s + t = 1$ and $0 \leq r, s, t \leq 1$.
Eq. (6.32) can be written as

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{bmatrix} \begin{bmatrix} r \\ s \\ t \end{bmatrix}$$

and is solved using Cramer's rule with

$$r = \frac{\begin{vmatrix} \mathbf{p} & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}} \quad s = \frac{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p} & \mathbf{p}_3 \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}} \quad t = \frac{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p} \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}}. \tag{6.33}$$

If r, s, and t satisfy the above constraints, then $\mathbf{p}$ is inside the triangle, otherwise it is outside the triangle's boundary. We will test these formulas in Section 6.15.

6.14 A sphere intersecting a plane

Detecting collisions between irregular objects is difficult. However, it is greatly simplified by enclosing an object within a tight-fitting sphere and calculating collisions between spheres. As part of this analysis, let's investigate the geometric relationship between a sphere and a plane. But apart from finding the conditions for collision, let's also compute the curve of intersection when a sphere intersects a plane.

We begin by positioning a sphere with radius R and centre $C(x_C, y_C, z_C)$ juxtaposed with a plane defined by

$$ax + by + cz = d$$

where

$$\hat{\mathbf{n}} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$$

and

$$\| \hat{\mathbf{n}} \| = 1$$

as shown in Fig. 6.20.

Now the shortest line between the sphere's centre C and the plane is when the line is perpendicular to the plane. Naturally, this line is parallel with the plane's normal $\hat{\mathbf{n}}$ and is represented vectorially by $\overrightarrow{CP}$. The length of $\overrightarrow{CP}$: $\left\| \overrightarrow{CP} \right\|$, informs us whether the sphere intersects, touches, or misses the plane:

$$\left\| \overrightarrow{CP} \right\| < R \text{ intersect condition}$$

$$\left\| \overrightarrow{CP} \right\| = R \text{ touch condition}$$

$$\left\| \overrightarrow{CP} \right\| > R \text{ miss condition.}$$

As $\| \hat{\mathbf{n}} \| = 1$, d represents the perpendicular distance from the origin to the plane.

Therefore, we can state that

$$\overrightarrow{CP} = \lambda \hat{\mathbf{n}} \tag{6.34}$$

where λ is a scalar.

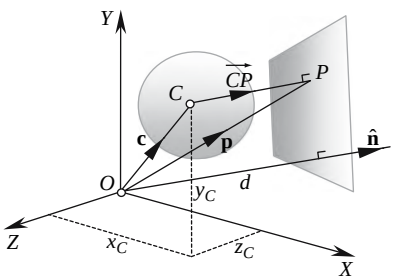


FIGURE 6.20.

The position vector $\mathbf{p}$ is defined as

$$\mathbf{p} = \mathbf{c} + \overrightarrow{CP} = \mathbf{c} + \lambda \hat{\mathbf{n}} \tag{6.35}$$

but as

$$\hat{\mathbf{n}} \cdot \mathbf{p} = d$$

$$\hat{\mathbf{n}} \cdot \mathbf{p} = \hat{\mathbf{n}} \cdot \mathbf{c} + \lambda \hat{\mathbf{n}} \cdot \hat{\mathbf{n}} = d$$

from which

$$\lambda = d - \hat{\mathbf{n}} \cdot \mathbf{c}. \tag{6.36}$$

But from Eq. (6.34), we see that

$$\left\| \overrightarrow{CP} \right\| = |\lambda|$$

and by finding $|\lambda|$, we find $\left\| \overrightarrow{CP} \right\|$ and the problem is solved.

Let's illustrate this technique with the following example: a plane is defined by

$$\frac{\sqrt{3}}{3}x + \frac{\sqrt{3}}{3}y + \frac{\sqrt{3}}{3}z = 10$$

and as its normal vector is a unit vector, the perpendicular distance from the origin to the plane is 10, as shown in Fig. 6.21.

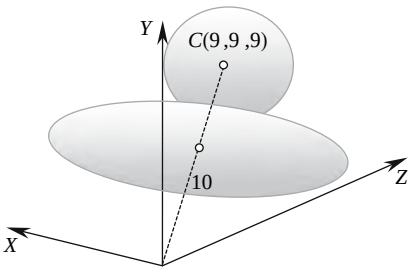


FIGURE 6.21.

The sphere's radius is 5 and its centre is located at (9, 9, 9), which makes it $\sqrt{243} = 15.588$ from the origin. Obviously, in this simple example we know that the sphere and plane do not intersect.

Using Eq. (6.36), we find that

$$|\lambda| = |d - \hat{\mathbf{n}} \cdot \mathbf{c}|$$

where

$$d = 10, \quad \hat{\mathbf{n}} = \frac{\sqrt{3}}{3}\mathbf{i} + \frac{\sqrt{3}}{3}\mathbf{j} + \frac{\sqrt{3}}{3}\mathbf{k}, \quad \mathbf{c} = 9\mathbf{i} + 9\mathbf{j} + 9\mathbf{k}$$

then

$$|\lambda| = \left| 10 - 9\frac{\sqrt{3}}{3} - 9\frac{\sqrt{3}}{3} - 9\frac{\sqrt{3}}{3} \right| = |-5.588|$$

$$|\lambda| = 5.588$$

and the sphere misses the plane because $|\lambda| > 5$.

If the plane is represented parametrically as shown in Fig. 6.22, then using two vectors **a** and **b** gives

$$\mathbf{p} = \mathbf{t} + \alpha\mathbf{a} + \beta\mathbf{b}$$

where α and β are scalars, we can proceed as follows.

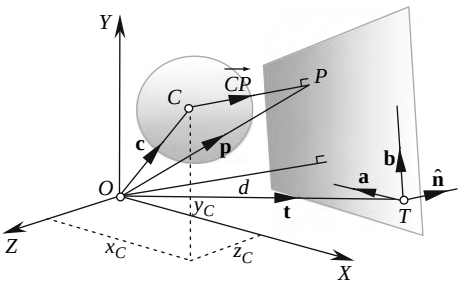


FIGURE 6.22.

We find the plane's normal vector **n** using the cross product

$$\mathbf{n} = \mathbf{a} \times \mathbf{b}$$

from which

$$\hat{\mathbf{n}} = \frac{\mathbf{n}}{\|\mathbf{n}\|}.$$

Therefore,

$$\hat{\mathbf{n}} \cdot \mathbf{t} = d$$

which can be substituted in Eq. (6.36):

$$\lambda = \hat{\mathbf{n}} \cdot \mathbf{t} - \hat{\mathbf{n}} \cdot \mathbf{c}$$
$$\lambda = \hat{\mathbf{n}} \cdot (\mathbf{t} - \mathbf{c})$$
$$\lambda = \frac{\mathbf{n}}{\|\mathbf{n}\|} \cdot (\mathbf{t} - \mathbf{c})$$

and

$$|\lambda| = \left| \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|} \cdot (\mathbf{t} - \mathbf{c}) \right|.$$

Let's test this technique with a simple example shown in Fig. 6.23.

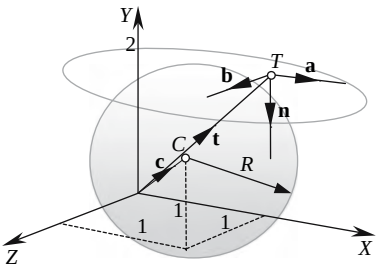


FIGURE 6.23.

The plane's equation is

$$\mathbf{p} = \mathbf{t} + \alpha\mathbf{a} + \beta\mathbf{b}$$

where

$$\mathbf{a} = \mathbf{i}, \quad \mathbf{b} = \mathbf{k}, \quad \mathbf{t} = \mathbf{i} + 2\mathbf{j}, \quad \mathbf{c} = \mathbf{i} + \mathbf{j} + \mathbf{k}, \quad R = 1.$$

Calculate $\mathbf{n}$:

$$\mathbf{n} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -\mathbf{j}$$

and

$$\hat{\mathbf{n}} = -\mathbf{j}.$$

Therefore,

$$|\lambda| = |(-\mathbf{j}) \cdot (\mathbf{i} + 2\mathbf{j} - \mathbf{i} - \mathbf{j} - \mathbf{k})|$$

$$|\lambda| = 1.$$

As $|\lambda| = R = 1$, the sphere touches the plane.

The parametric plane equation lends itself to calculating the circle of intersection when the sphere intersects the plane, so let's calculate the radius r of the intersecting circle. Figure 6.24(a) shows a cross section of the relevant geometry from which, using the theorem of Pythagoras,

$$R^2 = r^2 + (CP)^2,$$

but as

$$CP = |\lambda|$$

$$r = \sqrt{R^2 - \lambda^2}. \tag{6.37}$$

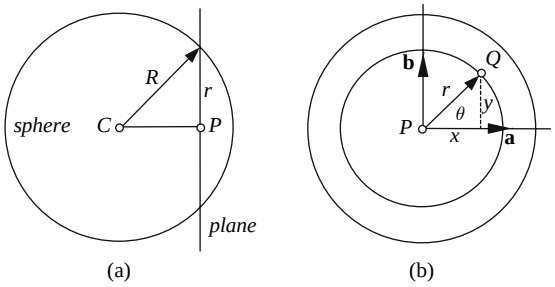


FIGURE 6.24.

By making **a** and **b** orthogonal and using the parametric equation of a circle:

$$x = r\cos\theta$$

$$y = r\sin\theta$$

as shown in Fig. 6.24(b), we can define a position vector **q** pointing to any point Q on the circle of intersection.

We already know that

$$\mathbf{p} = \mathbf{c} + \lambda\hat{\mathbf{n}}$$

where

$$\lambda = d - \hat{\mathbf{n}} \cdot \mathbf{c}$$

Therefore,

$$\mathbf{q} = \mathbf{p} + r\cos\theta\mathbf{a} + r\sin\theta\mathbf{b}$$

where

$$0° \leq \theta \leq 360°.$$

6.15 A sphere touching a triangle

Having seen how simple it is to detect a sphere touching or intersecting a plane, let's investigate the problem of a sphere touching a triangle. Basically, we need to follow three steps:

1. Derive the triangle's plane equation, probably using the cross product of two vectors taken from the triangle's edges.
2. Determine whether the sphere touches the triangle's plane.
3. If a touch condition occurs, discover if the touch point is inside the triangle.

To illustrate how the above steps are implemented, let's investigate a simple example.

Figure 6.25 shows a triangle $\Delta P_1P_2P_3$ located near the origin and a sphere of radius $R = 1$ with centre $C(1,\ 1,\ 2)$.

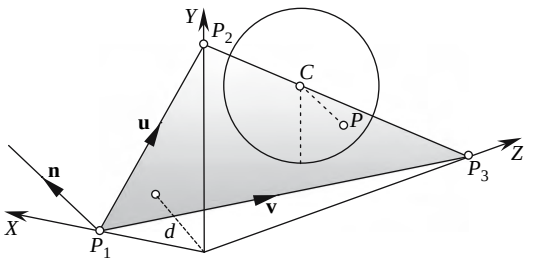


FIGURE 6.25.

We extract two vectors from the triangle

$$\mathbf{u} = \overrightarrow{P_1P_2} \quad \text{and} \quad \mathbf{v} = \overrightarrow{P_1P_3}$$

where

$$P_1 = (1,\ 0,\ 0) \quad P_2 = (0,\ 2,\ 0) \quad P_3 = (0,\ 0,\ 3)\,.$$

Therefore,

$$\mathbf{u} = -\mathbf{i} + 2\mathbf{j} \quad \text{and} \quad \mathbf{v} = -\mathbf{i} + 3\mathbf{k}.$$

Calculate $\mathbf{n}$:

$$\mathbf{n} = \mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 2 & 0 \\ -1 & 0 & 3 \end{vmatrix} = 6\mathbf{i} + 3\mathbf{j} + 2\mathbf{k}$$

$$\hat{\mathbf{n}} = \frac{6}{7}\mathbf{i} + \frac{3}{7}\mathbf{j} + \frac{2}{7}\mathbf{k}$$

and

$$\frac{6}{7}x + \frac{3}{7}y + \frac{2}{7}z = d \tag{6.38}$$

where d is the perpendicular distance from the origin to the plane. We can find d by substituting P_1 in Eq. (6.38):

$$d = \frac{6}{7}.$$

Using Eq. 6.36, where $|\,\lambda\,| = |\,d - \hat{\mathbf{n}} \cdot \mathbf{c}\,|$, we have

$$|\,\lambda\,| = \left| \frac{6}{7} - \left(\frac{6}{7}\mathbf{i} + \frac{3}{7}\mathbf{j} + \frac{2}{7}\mathbf{k} \right) \cdot (\mathbf{i} + \mathbf{j} + 2\mathbf{k}) \right|$$

$$|\,\lambda\,| = \left| \frac{6}{7} - \frac{13}{7} \right| = |\,-1\,|$$

$$|\,\lambda\,| = 1$$

which equals R and a touch condition occurs.

The negative value of λ informs us that the line connecting the sphere's centre C to the point P on the plane is against the direction of $\hat{\mathbf{n}}$.

Using Eq. (6.35), we can compute the touch point P:

$$\mathbf{p} = \mathbf{c} + \lambda\hat{\mathbf{n}} = (\mathbf{i}+\mathbf{j}+2\mathbf{k}) - \left(\frac{6}{7}\mathbf{i}+\frac{3}{7}\mathbf{j}+\frac{2}{7}\mathbf{k}\right)$$

and

$$P = \left(\frac{1}{7},\ \frac{4}{7},\ \frac{12}{7}\right).$$

Using Eq. (6.33), we can confirm whether this point is inside the triangle:

$$r = \frac{\begin{vmatrix} \mathbf{p} & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}} = \frac{\begin{vmatrix} \frac{1}{7} & 0 & 0 \\ \frac{4}{7} & 2 & 0 \\ \frac{12}{7} & 0 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix}} = \frac{\frac{6}{7}}{6} = \frac{1}{7}$$

$$s = \frac{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p} & \mathbf{p}_3 \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & \frac{1}{7} & 0 \\ 0 & \frac{4}{7} & 0 \\ 0 & \frac{12}{7} & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix}} = \frac{\frac{12}{7}}{6} = \frac{2}{7}$$

$$t = \frac{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p} \end{vmatrix}}{\begin{vmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 \end{vmatrix}} = \frac{\begin{vmatrix} 1 & 0 & \frac{1}{7} \\ 0 & 2 & \frac{4}{7} \\ 0 & 0 & \frac{12}{7} \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{vmatrix}} = \frac{\frac{24}{7}}{6} = \frac{4}{7}.$$

Note that $r+s+t=1$ and they are all positive and less than 1, which means that the point $\left(\frac{1}{7},\ \frac{4}{7},\ \frac{12}{7}\right)$ is inside the triangle, which means that the sphere touches the triangle within its boundary.

6.16 Two intersecting planes

Hopefully, it is obvious that two planes give rise to a straight line at their intersection. We can discover the vector representing this line by exploiting the fact that the intersecting line is perpendicular to the normal vectors associated with the planes. Therefore, the cross product of the plane's normal vectors reveals the vector, the direction, of which, is determined by the order of the normal vectors. Figure 6.26 illustrates such a scenario.

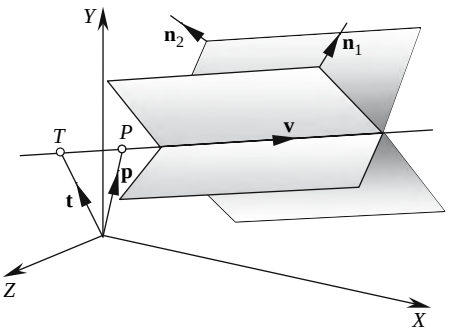


FIGURE 6.26.

To derive a parametric line equation for the intersection, we require the coordinates of a point on the line. It is convenient to select a point whose position vector $\mathbf{t}$ is perpendicular to $\mathbf{v}$.

We begin by defining the two plane equations

$$a_1x + b_1y + c_1z = d_1$$
$$a_2x + b_2y + c_2z = d_2$$

where

$$\mathbf{n}_1 = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$$
$$\mathbf{n}_2 = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}.$$

Therefore,

$$\mathbf{n}_1 \times \mathbf{n}_2 = \mathbf{v}$$

where $\mathbf{v}$ is the vector representing the line of intersection.

The line equation is

$$\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$$

where

λ is a scalar.

As $T(x_T, y_T, z_T)$ must satisfy both plane equations,

$$\mathbf{n}_1 \cdot \mathbf{t} = d_1 \text{ and } \mathbf{n}_2 \cdot \mathbf{t} = d_2,$$

Furthermore, as $\mathbf{v}$ and $\mathbf{t}$ are orthogonal,

$$\mathbf{v} \cdot \mathbf{t} = 0.$$

These three vector equations can be combined and written as

$$\begin{bmatrix} d_1 \\ d_2 \\ 0 \end{bmatrix} = \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_v & b_v & c_v \end{bmatrix} \cdot \begin{bmatrix} x_T \\ y_T \\ z_T \end{bmatrix}$$

and solved using Cramer's rule, where

$$x_T = \frac{\begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ 0 & b_v & c_v \end{vmatrix}}{DET} \quad y_T = \frac{\begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_v & 0 & c_v \end{vmatrix}}{DET} \quad z_T = \frac{\begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_v & b_v & 0 \end{vmatrix}}{DET}$$

$$DET = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_v & b_v & c_v \end{vmatrix}$$

and

$$\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix}.$$

To illustrate this technique, let's consider a scenario whose outcome can be predicted as shown in Fig. 6.27, where two planes intersect along the line $\mathbf{p} = \mathbf{t} + \lambda\mathbf{v}$, where $\mathbf{t} = \mathbf{j}$ and $\mathbf{v} = \mathbf{i}$.
The plane equations are

$$y - z = 1 \quad \text{and} \quad y + z = 1.$$

Therefore,

$$\mathbf{n}_1 = \mathbf{j} - \mathbf{k} \quad \text{and} \quad \mathbf{n}_2 = \mathbf{j} + \mathbf{k}$$

and

$$\mathbf{v} = \mathbf{n}_1 \times \mathbf{n}_2 = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{vmatrix} = 2\mathbf{i}$$

$$DET = \begin{vmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ 2 & 0 & 0 \end{vmatrix} = 4$$

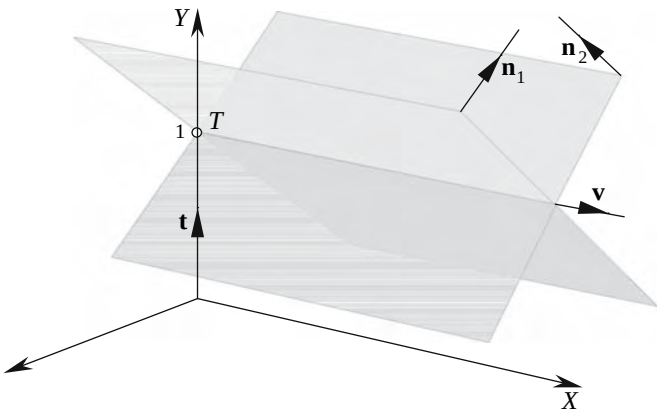


FIGURE 6.27.

$$x_T = \frac{1}{4}\begin{vmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{vmatrix} = 0$$

$$y_T = \frac{1}{4}\begin{vmatrix} 0 & 1 & -1 \\ 0 & 1 & 1 \\ 2 & 0 & 0 \end{vmatrix} = \frac{4}{4} = 1$$

$$z_T = \frac{1}{4}\begin{vmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 2 & 0 & 0 \end{vmatrix} = 0.$$

Therefore, the line equation is

$$\mathbf{p} = \mathbf{j} + \lambda\mathbf{i}.$$

Remember that the direction of $\mathbf{v}$ could have been reversed if $\mathbf{n}_1$ and $\mathbf{n}_2$ had been reversed in the cross-product calculation.

This example concludes this chapter on intersections. Hopefully, the reader will feel confident to undertake other intersection problems by implementing some of the above strategies.

7 Rotating Vectors

7.1 Introduction

In this chapter we develop a transform to rotate a position vector about an arbitrary axis and then we examine how quaternions perform a similar function, but using a totally different approach.

Points in R^3 are readily rotated about one or more Cartesian axes using Euler transforms, which are, unfortunately, prone to gimbal lock, where a degree of rotational freedom is lost under certain circumstances. However, in the 19th century, quaternions were discovered, which can rotate vectors in R^3 without gimbal lock. But before examining quaternions, let's develop a transform to rotate a vector about an arbitrary axis.

7.2 Rotating a vector about an arbitrary axis

This analysis assumes a position vector $\mathbf{p}$ with its associated point P, which is to be rotated about an axis $\hat{\mathbf{n}}$ that intersects the origin. Such a scheme is shown in Fig. 7.1, where Fig. 7.1(a) shows the overall scenario, Fig. 7.1(b) illustrates a side elevation, and Fig. 7.1(c) illustrates a plan elevation.

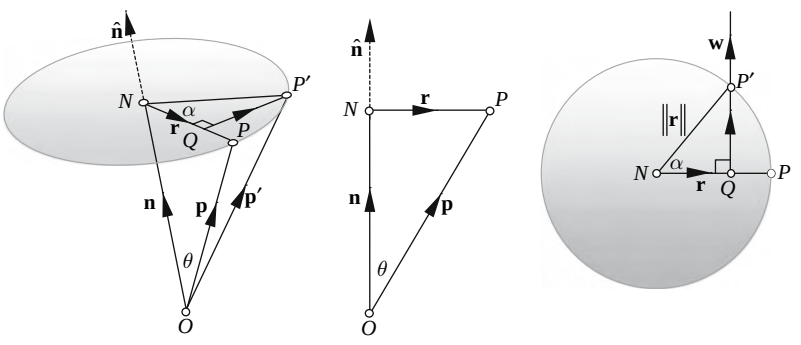


Figure 7.1.

In Fig. 7.1,

$\hat{\mathbf{n}} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ is the unit vector representing the axis of rotation,
$P(x_p, y_p, z_p)$ is the point to be rotated by angle α,
$P'(x'_p, y'_p, z'_p)$ is the rotated point,
$\mathbf{p}$ and $\mathbf{p}'$ are position vectors for P and P', respectively,
O is the origin.

The objective is to derive a vector equation describing $\mathbf{p}'$ in terms of $\hat{\mathbf{n}}$, $\mathbf{p}$, and α.

So let's start by defining $\mathbf{p}'$ in terms of vectors associated with the construction lines shown in Fig. 7.1(a):

$$\mathbf{p}' = \overrightarrow{ON} + \overrightarrow{NQ} + \overrightarrow{QP'}.$$

We now define each of these vectors in turn.

To find $\overrightarrow{ON}$, we start by taking the dot product of $\hat{\mathbf{n}}$ and $\mathbf{p}$:

$$\hat{\mathbf{n}} \cdot \mathbf{p} = \| \hat{\mathbf{n}} \| \| \mathbf{p} \| \cos\theta = \| \mathbf{p} \| \cos\theta = ON.$$

Therefore,

$$\overrightarrow{ON} = (\hat{\mathbf{n}} \cdot \mathbf{p})\,\hat{\mathbf{n}} = \mathbf{n}. \tag{7.1}$$

We use the right-angle triangle $\Delta NQP'$ to find $\overrightarrow{NQ}$:

$$\overrightarrow{NQ} = \frac{NQ}{NP}\mathbf{r} = \frac{NQ}{NP'}\mathbf{r} = \cos\alpha\,\mathbf{r}. \tag{7.2}$$

To eliminate $\mathbf{r}$ in Eq. (7.2), we construct an equation defining $\mathbf{r}$ in terms of known vectors:

$$\mathbf{r} = \mathbf{p} - \mathbf{n}. \tag{7.3}$$

Substituting Eq. (7.1) in Eq. (7.3), we obtain

$$\mathbf{r} = \mathbf{p} - (\hat{\mathbf{n}} \cdot \mathbf{p})\,\hat{\mathbf{n}}$$

and

$$\overrightarrow{NQ} = [\mathbf{p} - (\hat{\mathbf{n}} \cdot \mathbf{p})\,\hat{\mathbf{n}}]\cos\alpha.$$

To find $\overrightarrow{QP'}$, we exploit the fact that $\mathbf{w}$ [Fig. 7.1(c)] is perpendicular to the plane containing $\hat{\mathbf{n}}$ and $\mathbf{p}$, such that

$$\hat{\mathbf{n}} \times \mathbf{p} = \mathbf{w}$$

where

$$\| \mathbf{w} \| = \| \hat{\mathbf{n}} \| \| \mathbf{p} \| \sin\theta = \| \mathbf{p} \| \sin\theta. \tag{7.4}$$

But from Fig. 7.1(b), we have

$$\| \mathbf{r} \| = \| \mathbf{p} \| \sin\theta. \tag{7.5}$$

Therefore Eq. (7.4) equals Eq. (7.5) and

$$\| \mathbf{w} \| = \| \mathbf{r} \|.$$

From Fig. 7.1(c), we know that

$$\frac{QP'}{NP'} = \frac{QP'}{\| \mathbf{r} \|} = \frac{QP'}{\| \mathbf{w} \|} = \sin\alpha$$

Therefore,

$$\overrightarrow{QP'} = \mathbf{w}\sin\alpha = (\hat{\mathbf{n}} \times \mathbf{p})\sin\alpha$$

and then

$$\mathbf{p}' = (\hat{\mathbf{n}} \cdot \mathbf{p})\,\hat{\mathbf{n}} + [\mathbf{p} - (\hat{\mathbf{n}} \cdot \mathbf{p})\,\hat{\mathbf{n}}]\cos\alpha + (\hat{\mathbf{n}} \times \mathbf{p})\sin\alpha$$

and

$$\mathbf{p}' = \mathbf{p}\cos\alpha + (\hat{\mathbf{n}} \cdot \mathbf{p})\,(1 - \cos\alpha)\,\hat{\mathbf{n}} + (\hat{\mathbf{n}} \times \mathbf{p})\sin\alpha.$$

We make the following substitution:

$$K = 1 - \cos\alpha$$

then

$$\mathbf{p}' = \mathbf{p}\cos\alpha + (\hat{\mathbf{n}} \cdot \mathbf{p})\,K\hat{\mathbf{n}} + (\hat{\mathbf{n}} \times \mathbf{p})\sin\alpha$$

and

$$\begin{aligned}
\mathbf{p}' &= \left(x_p\mathbf{i} + y_p\mathbf{j} + z_p\mathbf{k}\right)\cos\alpha + \left(ax_p + by_p + cz_p\right)K\,(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \\
&\quad + \left[\left(bz_p - cy_p\right)\,\mathbf{i} + \left(cx_p - az_p\right)\,\mathbf{j} + \left(ay_p - bx_p\right)\,\mathbf{k}\right]\sin\alpha \\
\mathbf{p}' &= \left[x_p\cos\alpha + a\left(ax_p + by_p + cz_p\right)K + \left(bz_p - cy_p\right)\sin\alpha\right]\,\mathbf{i} \\
&\quad + \left[y_p\cos\alpha + b\left(ax_p + by_p + cz_p\right)K + \left(cx_p - az_p\right)\sin\alpha\right]\,\mathbf{j} \\
&\quad + \left[z_p\cos\alpha + c\left(ax_p + by_p + cz_p\right)K + \left(ay_p - bx_p\right)\sin\alpha\right]\,\mathbf{k} \\
\mathbf{p}' &= \left[x_p\left(a^2K + \cos\alpha\right) + y_p\left(abK - c\sin\alpha\right) + z_p\left(acK + b\sin\alpha\right)\right]\,\mathbf{i} \\
&\quad + \left[x_p\left(abK + c\sin\alpha\right) + y_p\left(b^2K + \cos\alpha\right) + z_p\left(bcK - a\sin\alpha\right)\right]\,\mathbf{j} \\
&\quad + \left[x_p\left(acK - b\sin\alpha\right) + y_p\left(bcK + a\sin\alpha\right) + z_p\left(c^2K + \cos\alpha\right)\right]\,\mathbf{k}
\end{aligned}$$

and the transform becomes

$$\begin{bmatrix} x'_p \\ y'_p \\ z'_p \end{bmatrix} = \begin{bmatrix} a^2K + \cos\alpha & abK - c\sin\alpha & acK + b\sin\alpha \\ abK + c\sin\alpha & b^2K + \cos\alpha & bcK - a\sin\alpha \\ acK - b\sin\alpha & bcK + a\sin\alpha & c^2K + \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix}$$

where $K = 1 - \cos\alpha$.
To test this transform, let's perform a simple rotation such as rotating the point $P\,(1, 1, 0)$ $90°$ about the x-axis $\hat{\mathbf{n}} = \mathbf{i}$. With reference to Fig. 7.2, it is obvious that the answer is $P'\,(1, 0, 1)$.

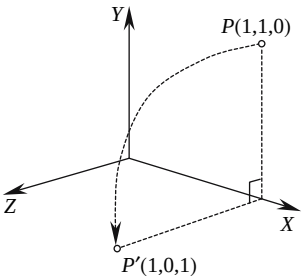


FIGURE 7.2.

If $\theta = 90°$, then $K = 1 - \cos 90° = 1$. Also, if the axis of rotation is $\hat{\mathbf{n}} = \mathbf{i}$, then $a = 1$, $b = 0$, and $c = 0$. Therefore, the transform becomes

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

which is correct.

7.3 Complex numbers

Complex numbers were discovered in the 16th century but were not fully embraced by mathematicians, who tended to endorse their "imaginary" associations. Eventually, in the early 19th century, Carl Friedrich Gauss [1777–1855] showed that complex numbers had a geometric interpretation, and the mathematical landscape was prepared for a fertile period of discovery.

Prior to the discovery of complex numbers, it was difficult to manipulate the square root of a negative number. However, with the invention of $\mathbf{i}$, which could stand in place of $\sqrt{-1}$, the imaginary world of complex numbers came into being.

By definition, a complex number z has the form

$$z = a + \mathbf{i}b \text{ or } z = a + b\mathbf{i}$$

where a and b are real quantities and $\mathbf{i}^2 = -1$. The position of $\mathbf{i}$ is not important, and in this text it is placed after the scalar. An example is $2 + 3\mathbf{i}$, where the real part is 2 and the imaginary

part is 3**i**. There is no requirement for a complex number to have a real part, in fact, even the imaginary part could be zero, which implies that the set of real numbers $\mathbb{R}$. is a subset of the complex number set $\mathbb{Z}$.

What is interesting is that even if we ignore what complex numbers are or what they can do, they are extremely well behaved when subjected to the normal laws of algebra. For example, we can add, subtract, multiply, divide, raise to a power, take logs, etc. of a complex number.

7.3.1 Complex number operations

Briefly, here are the rules associated with complex numbers:

Equivalence: Two complex numbers are equal only if both real and imaginary parts are the same. For example, if

$$a + b\mathbf{i} = r + s\mathbf{i}$$

then

$$a = r \quad \text{and} \quad b = s.$$

Addition and subtraction: To add or subtract two complex numbers, the addition or subtraction is performed individually on the real and imaginary parts:

$$(a + b\mathbf{i}) \pm (r + s\mathbf{i}) = (a \pm r) + (b \pm s)\,\mathbf{i}.$$

For example,

$$(12 + 3\mathbf{i}) - (9 - 4\mathbf{i}) = 3 + 7\mathbf{i}.$$

Multiplication: To multiply two complex numbers together, we expand the multiplication using the conventional rules of algebra:

$$(a + b\mathbf{i}) \times (r + s\mathbf{i}) = ar + as\mathbf{i} + br\mathbf{i} + bs\mathbf{i}^2.$$

Collecting like terms, we obtain

$$(a + b\mathbf{i}) \times (r + s\mathbf{i}) = (ar - bs) + (as + br)\,\mathbf{i}.$$

Note that the $\mathbf{i}^2$ term makes the product *bs* negative.
For example,

$$(12 + 3\mathbf{i}) \times (9 - 4\mathbf{i}) = 108 + 27\mathbf{i} - 48\mathbf{i} - 12\mathbf{i}^2 = 120 - 21\mathbf{i}.$$

Division: To divide a complex number by another, we invoke the following subterfuge:

$$\frac{a + b\mathbf{i}}{r + s\mathbf{i}} = \frac{(a + b\mathbf{i})}{(r + s\mathbf{i})}\frac{(r - s\mathbf{i})}{(r - s\mathbf{i})} = \frac{ar - as\mathbf{i} + br\mathbf{i} - bs\mathbf{i}^2}{r^2 - rs\mathbf{i} + rs\mathbf{i} - s^2\mathbf{i}^2} = \frac{(ar + bs) + (br - as)\,\mathbf{i}}{r^2 + s^2}$$

$$\frac{a + b\mathbf{i}}{r + s\mathbf{i}} = \frac{(ar + bs)}{r^2 + s^2} + \frac{(br - as)}{r^2 + s^2}\mathbf{i}.$$

For example,

$$\frac{(12+3\mathbf{i})}{(9-4\mathbf{i})}=\frac{(12+3\mathbf{i})}{(9-4\mathbf{i})}\frac{(9+4\mathbf{i})}{(9+4\mathbf{i})}=\frac{108-12}{81+16}+\frac{27+48}{81+16}\mathbf{i}=\frac{96}{97}+\frac{75}{97}\mathbf{i}.$$

7.3.2 Magnitude of a complex number

If $z=a+b\mathbf{i}$, then its magnitude or absolute value is defined as

$$|z|=\sqrt{a^2+b^2}.$$

For example,

$$|12+3\mathbf{i}|=\sqrt{144+9}=\sqrt{153}.$$

7.3.3 The complex conjugate

If $z=a+b\mathbf{i}$, then its complex conjugate is defined as

$$\bar{z}=a-b\mathbf{i}$$

and was used above to simplify the division of two complex numbers.

Just by defining the complex conjugate produces some interesting identities. For example,

$$z\bar{z}=(a+b\mathbf{i})(a-b\mathbf{i})=a^2+b^2$$

$$z+\bar{z}=(a+b\mathbf{i})+(a-b\mathbf{i})=2a$$

$$\overline{z_1}+\overline{z_2}=(a_1-b_1\mathbf{i})+(a_2-b_2\mathbf{i})=(a_1+a_2)-(b_1+b_2)\mathbf{i}=\overline{z_1+z_2}$$

$$\overline{z_1}\,\overline{z_2}=(a_1-b_1\mathbf{i})(a_2-b_2\mathbf{i})=(a_1a_2-b_1b_2)-(a_1b_2+b_1a_2)\mathbf{i}=\overline{z_1z_2}$$

and

$$z\bar{z}=|z|^2=a^2+b^2 \tag{7.6}$$

or

$$|z|=\sqrt{z\bar{z}}.$$

Furthermore, using Eq. (7.6), we find that

$$|z|^2=z\bar{z}$$

Therefore,

$$|z_1z_2|^2=(z_1z_2)\,\overline{(z_1z_2)}$$

$$|z_1z_2|^2=z_1z_2\,\overline{z_1}\,\overline{z_2}$$

$$|z_1z_2|^2=z_1\overline{z_1}z_2\,\overline{z_2}=|z_1|^2|z_2|^2$$

or

$$|z_1z_2|=|z_1||z_2|.$$

7.3.4 Graphical interpretation of complex numbers

In 1806 Jean Robert Argand [1768–1822] showed that complex numbers had a useful geometrical interpretation and could represent points in $\mathbb{R}^2$. The breakthrough was to represent a complex number as the coordinates of a point. But instead of using x-and y-axes, we employ real and imaginary axes, as shown in Fig. 7.3.

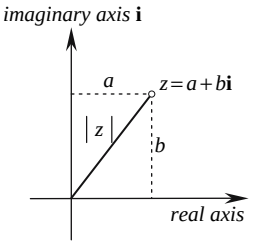


FIGURE 7.3.

7.3.5 i as a rotator

Once there was a geometric interpretation for complex numbers, it must have been tempting to investigate the geometric impact arithmetic operations would have upon them. So let's explore the geometric effect of repeatedly multiplying a complex number by **i**.

If the starting number is $z = a$, i.e., no imaginary component, then by multiplying z by **i** gives us $a\mathbf{i}$. Multiplying again by **i**, we obtain $-a$. The next two multiplications produce $-a\mathbf{i}$ and a, respectively. The four multiplications have effectively rotated the complex number 360°, as shown in Fig. 7.4.

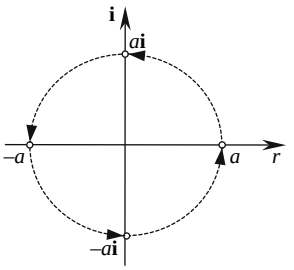


FIGURE 7.4.

This rotational ability applies to any complex number. For example, if the starting complex number is $2+\mathbf{i}$, and we repeatedly multiply by **i**, we obtain

$$-1+2\mathbf{i}, \quad -2-\mathbf{i}, \quad 1-2\mathbf{i}, \quad 2+\mathbf{i}$$

whose positions are shown in Fig. 7.5.

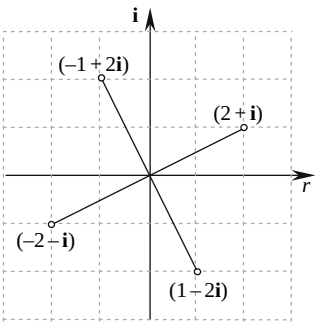


FIGURE 7.5.

7.3.6 Unifying *e*, *i*, sin, and cos

e^x, $\sin x$, $\cos x$ can be defined as the infinite series

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots$$
$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \cdots .$$

But what happens if we make x a complex number in e^x?

$$e^{x\mathbf{i}} = 1 + \frac{x\mathbf{i}}{1!} + \frac{x^2\mathbf{i}^2}{2!} + \frac{x^3\mathbf{i}^3}{3!} + \cdots$$
$$e^{x\mathbf{i}} = 1 + \frac{x\mathbf{i}}{1!} - \frac{x^2}{2!} - \frac{x^3\mathbf{i}}{3!} + \frac{x^4}{4!} + \frac{x^5\mathbf{i}}{5!} \cdots .$$

Collecting real and imaginary terms gives us

$$e^{x\mathbf{i}} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} \cdots + \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \cdots \right) \mathbf{i}$$

which we recognise as the sin and cos functions. Therefore,

$$e^{x\mathbf{i}} = \cos x + \sin x\mathbf{i}. \tag{7.7}$$

When $x = \pi$, we obtain

$$e^{\pi\mathbf{i}} = \cos \pi + \sin \pi\mathbf{i}$$
$$e^{\pi\mathbf{i}} = -1 \tag{7.8}$$

which is the beautiful relationship discovered by Euler.

Now what is really strange about $\mathbf{i}$ is that it can even be raised to itself: $\mathbf{i}^{\mathbf{i}}$. To discover its value, we rearrange Eq. (7.8):

$$-1 = e^{\pi \mathbf{i}}.$$

Now if $e^{\pi \mathbf{i}}$ is equivalent to a rotation of π radians, $e^{\frac{\pi}{2}\mathbf{i}}$ must be equivalent to a rotation of 90°. Therefore,

$$\mathbf{i} = e^{\frac{\pi}{2}\mathbf{i}}.$$

Now we introduce $\mathbf{i}^{\mathbf{i}}$:

$$\mathbf{i}^{\mathbf{i}} = \left(e^{\frac{\pi}{2}\mathbf{i}}\right)^{\mathbf{i}}.$$

Therefore,

$$\mathbf{i}^{\mathbf{i}} = e^{-\frac{\pi}{2}} = 0.2078795\ldots.$$

7.3.7 Complex numbers as rotators

Equation (7.7) behaves like a rotator, and the RHS $\cos x + \mathbf{i}\sin x$ is a complex number that rotates another complex number by x radians. For example, if $x = 45°$, then

$$\cos x + \sin x\mathbf{i} = 0.707 + 0.707\mathbf{i}.$$

If another complex number is multiplied by $0.707 + 0.707\mathbf{i}$, it is rotated by 45°. For example, to rotate $2 + 2\mathbf{i}$ by 45°, we compute

$$(0.707 + 0.707\mathbf{i})(2 + 2\mathbf{i}) = 1.414 + 1.414\mathbf{i} + 1.414\mathbf{i} + 1.414\mathbf{i}^2 = 2.828\mathbf{i}$$

which is correct. The magnitude of $2 + 2\mathbf{i}$ is

$$|2 + 2\mathbf{i}| = 2.828$$

and is preserved after the rotation.

We can now state that any vector $\mathbf{v} = \begin{bmatrix} x & y \end{bmatrix}^T$ is rotated by an angle θ using

$$z = (\cos\theta + \sin\theta\mathbf{i})(x + y\mathbf{i}) \tag{7.9}$$

where the real and imaginary components of z are the x and y components, respectively, of the rotated vector.

But before we move on, let's take one last look at Eq. (7.9). Expanding the terms, we obtain

$$z = x\cos\theta + x\sin\theta\mathbf{i} + y\cos\theta\mathbf{i} + y\sin\theta\mathbf{i}^2$$

$$z = x\cos\theta - y\sin\theta + (x\sin\theta + y\cos\theta)\,\mathbf{i}$$

or in matrix form

$$z = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta\mathbf{i} & \cos\theta\mathbf{i} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

which is the rotation matrix in a complex disguise!

7.4 Quaternions

Knowing that a complex number rotates a position vector in R^2, it then seems natural that there should be a R^3 equivalent, but the search for such a number proved extremely difficult. Nevertheless, after many years of toil, Sir William Rowan Hamilton discovered its form on October 16, 1843.

Prior to this discovery, Sir William had toyed with the idea that a hyper-complex number would take the form

$$z = a + b\mathbf{i} + c\mathbf{j}$$

where

$$\mathbf{i}^2 = \mathbf{j}^2 = -1.$$

However, a problem occurs when two such numbers are multiplied together:

$$(a_1 + b_1\mathbf{i} + c_1\mathbf{j})(a_2 + b_2\mathbf{i} + c_2\mathbf{j}) = a_1a_2 + a_1b_2\mathbf{i} + a_1c_2\mathbf{j} + b_1a_2\mathbf{i} + b_1b_2\mathbf{i}^2 + b_1c_2\mathbf{ij} + c_1a_2\mathbf{j} + c_1b_2\mathbf{ji} + c_1c_2\mathbf{j}^2$$

$$(a_1 + b_1\mathbf{i} + c_1\mathbf{j})(a_2 + b_2\mathbf{i} + c_2\mathbf{j}) = (a_1a_2 - b_1b_2 - c_1c_2) + (a_1b_2 + b_1a_2)\,\mathbf{i} + (a_1c_2 + c_1a_2)\,\mathbf{j} + b_1c_2\mathbf{ij} + c_1b_2\mathbf{ji}$$

The new number contains a real part, two imaginary parts, and two further parts containing the products **ij** and **ji**, for which no meaning could be found.

Sir William then tried the next obvious form:

$$z = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$$

and when two such numbers are multiplied together, they create a large number of terms:

$$\begin{aligned} q_1 &= (a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}) \\ q_2 &= (a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) \\ q_1q_2 =& a_1a_2 + a_1(b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) + a_2(b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}) \\ &+ b_1b_2\mathbf{i}^2 + b_1c_2\mathbf{ij} + b_1d_2\mathbf{ik} \\ &+ c_1b_2\mathbf{ji} + c_1c_2\mathbf{j}^2 + c_1d_2\mathbf{jk} \\ &+ d_1b_2\mathbf{ki} + d_1c_2\mathbf{kj} + d_1d_2\mathbf{k}^2 \end{aligned}$$

Invoking the rule $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$, we obtain

$$\begin{aligned} q_1q_2 = a_1a_2 - (b_1b_2 + c_1c_2 + d_1d_2) + a_1(b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}) + a_2(b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}) \\ + b_1c_2\mathbf{ij} + b_1d_2\mathbf{ik} + c_1b_2\mathbf{ji} + c_1d_2\mathbf{jk} + d_1b_2\mathbf{ki} + d_1c_2\mathbf{kj} \end{aligned} \tag{7.10}$$

The new number contains a real part, three imaginary parts, and six further parts containing the products **ij**, **ik**, **ji**, **jk**, **ki**, and **kj**, for which no obvious meaning could be found. And then on that fateful day in 1843, Hamilton thought of the extra rules:

$$\mathbf{ij} = \mathbf{k}, \mathbf{jk} = \mathbf{i}, \mathbf{ki} = \mathbf{j}, \mathbf{ji} = -\mathbf{k}, \mathbf{kj} = -\mathbf{i} \text{ and } \mathbf{ik} = -\mathbf{j}$$

which, when substituted in Eq. (7.10), produce

$$\begin{aligned} q_1 q_2 =& a_1 a_2 - (b_1 b_2 + c_1 c_2 + d_1 d_2) \\ &+ a_1 (b_2 \mathbf{i} + c_2 \mathbf{j} + d_2 \mathbf{k}) + a_2 (b_1 \mathbf{i} + c_1 \mathbf{j} + d_1 \mathbf{k}) \\ &+ (c_1 d_2 - d_1 c_2)\, \mathbf{i} + (d_1 b_2 - b_1 d_2)\, \mathbf{j} + (b_1 c_2 - c_1 b_2)\, \mathbf{k} \end{aligned} \tag{7.11}$$

Equation (7.11) has an obvious structure, which probably did not emerge immediately. However, with our knowledge of vectors, it is possible to recognise the dot product:

$$(b_1 b_2 + c_1 c_2 + d_1 d_2)$$

a vector multiplied by a scalar

$$a_1 (b_2 \mathbf{i} + c_2 \mathbf{j} + d_2 \mathbf{k})$$

and

$$a_2 (b_1 \mathbf{i} + c_1 \mathbf{j} + d_1 \mathbf{k})$$

and the cross product

$$(c_1 d_2 - d_1 c_2)\, \mathbf{i} + (d_1 b_2 - b_1 d_2)\, \mathbf{j} + (b_1 c_2 - c_1 b_2)\, \mathbf{k}.$$

Hamilton gave the name *quaternion* to this form and the name *vector* to its imaginary component. He had also given birth to non-commutative algebra, as the quaternion product is non-commutative, i.e., $\mathbf{ij} \neq \mathbf{ji}$. Hyper-complex numbers also became a new set for describing numbers. Today, the set of hyper-complex numbers embraces real numbers, which have a rank 1, complex numbers, which have a rank 2, and quaternions, which have a rank 4.

Thus, the product of two quaternions

$$q_1 = (s_1 + \mathbf{v}_1) \quad \text{and} \quad q_2 = (s_2 + \mathbf{v}_2)$$

equals

$$q_1 q_2 = (s_1 s_2 - v_1 \cdot v_2) + s_1 \mathbf{v}_2 + s_2 \mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2.$$

Hamilton was so pleased with his discovery that he immediately wrote to his mathematician, friend John Graves [1806–1870], who had encouraged him in the first place to pursue this avenue of research. By December of the same year, Graves had discovered the next algebra of rank 8, which he called *octaves*. He also proposed that other algebras might exist of the form "2^m-ions", where 2^0 are real numbers, 2^1 are complex numbers, 2^2 are quaternions, and 2^3 are octaves. He tried, in vain, to find an algebra of the form 2^4.

Unfortunately for Graves, his idea of octaves was published after Arthur Cayley [1821–1895] published his own discovery of *octonions* in 1845, and Cayley's publishing priority ensured that he was credited with their discovery. To learn more about quaternions and octonions, John Conway's and Derek Smith's book, *On Quaternions and Octonions*, is highly recommended [Conway, 2003].

Although some mathematicians of the day supported quaternions, the newly discovered vector had even more supporters and by 1900 vector analysis was the preferred analytical tool. By the 1960s quaternions were being used in flight simulators, and in the 1980s Shoemake [Shoemake, 1985] found a use for them in computer graphics. Today, they are used in computer animation systems and computer-game development systems.

7.5 Quaternions as rotators

Surely, if a complex number can rotate a vector in R^2, can a quaternion rotate a vector in R^3? Let's explore how this is possible.

In Section 7.3.5 we saw that a complex number can be used as a 2D rotator using

$$z = (\cos\theta + \sin\theta\mathbf{i})\,(x + y\mathbf{i})$$

and it is tempting to see whether a quaternion can be used in a similar fashion. This time we will use a vector

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

and a quaternion

$$q = \cos\theta + \sin\theta\mathbf{i}$$

and create the product

$$q\mathbf{v} = (\cos\theta + \sin\theta\mathbf{i})\,(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$$

which expands to

$$q\mathbf{v} = x\cos\theta\mathbf{i} + y\cos\theta\mathbf{j} + z\cos\theta\mathbf{k} + x\sin\theta\mathbf{i}^2 + y\sin\theta\mathbf{ij} + z\sin\theta\mathbf{ik}.$$

Invoking Hamilton's rules ($\mathbf{i}^2 = -1, \quad \mathbf{ij} = \mathbf{k}, \quad \mathbf{ik} = -\mathbf{j}$) gives

$$q\mathbf{v} = x\cos\theta\mathbf{i} + y\cos\theta\mathbf{j} + z\cos\theta\mathbf{k} - x\sin\theta + y\sin\theta\mathbf{k} - z\sin\theta\mathbf{j}.$$

Collecting like terms gives

$$q\mathbf{v} = -x\sin\theta + x\cos\theta\mathbf{i} + (y\cos\theta - z\sin\theta)\,\mathbf{j} + (z\cos\theta + y\sin\theta)\,\mathbf{k}.$$

Some subtle changes reveal

$$q\mathbf{v} = (x\cos\theta + x\sin\theta\mathbf{i})\,\mathbf{i} + (y\cos\theta - z\sin\theta)\,\mathbf{j} + (z\cos\theta + y\sin\theta)\,\mathbf{k}. \tag{7.12}$$

Equation (7.12) is tantalizingly close to a perfect rotation about the **i**-axis. For if we ignore the **i** term, the **j** and **k** terms contain a rotation transform for the y and z coordinates:

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}.$$

Obviously, the operation $q\mathbf{v}$ falls short of a pure rotation, but we could consider developing it to create one. Basically, we want to convert the **i** term to $x\mathbf{i}$, which can, perhaps, be achieved by extending $q\mathbf{v}$ to $q\mathbf{v}\bar{q}$, where $\bar{q}$ is another quaternion.

Let's make

$$\bar{q} = \cos\theta - \sin\theta\mathbf{i}$$

then

$$q\mathbf{v}\bar{q} = (x\cos\theta + x\sin\theta\mathbf{i})\,\mathbf{i}\,(\cos\theta - \sin\theta\mathbf{i}) + [(y\cos\theta - z\sin\theta)\,\mathbf{j} + (z\cos\theta + y\sin\theta)\,\mathbf{k}]\,(\cos\theta - \sin\theta\mathbf{i})$$

and

$$q\mathbf{v}\bar{q} = \left(x\cos^2\theta + x\sin^2\theta\right)\mathbf{i} + [(y\cos\theta - z\sin\theta)\,\mathbf{j} + (z\cos\theta + y\sin\theta)\,\mathbf{k}]\,(\cos\theta - \sin\theta\mathbf{i})\,. \quad (7.13)$$

Equation (7.13) then simplifies to

$$q\mathbf{v}\bar{q} = x\mathbf{i} + [(y\cos\theta - z\sin\theta)\,\mathbf{j} + (z\cos\theta + y\sin\theta)\,\mathbf{k}]\,(\cos\theta - \sin\theta\mathbf{i}) \quad (7.14)$$

which seems to have done the trick. Unfortunately, the remaining **j** and **k** terms have been modified, so let's expand Eq. (7.14):

$$q\mathbf{v}\bar{q} = x\mathbf{i} + (y\cos\theta\mathbf{j} - z\sin\theta\mathbf{j})\,(\cos\theta - \sin\theta\mathbf{i}) + (z\cos\theta\mathbf{k} + y\sin\theta\mathbf{k})\,(\cos\theta - \sin\theta\mathbf{i})\,,$$

$$q\mathbf{v}\bar{q} = x\mathbf{i} + y\cos^2\theta\mathbf{j} + y\cos\theta\sin\theta\mathbf{k} - z\cos\theta\sin\theta\mathbf{j} - z\sin^2\theta\mathbf{k} + z\cos^2\theta\mathbf{k} - z\cos\theta\sin\theta\mathbf{j} + y\cos\theta\sin\theta\mathbf{k} - y\sin^2\theta\mathbf{j}.$$

Collecting like terms gives

$$q\mathbf{v}\bar{q} = x\mathbf{i} + \left(y\cos^2\theta - y\sin^2\theta - 2z\cos\theta\sin\theta\right)\mathbf{j} + \left(z\cos^2\theta - z\sin^2\theta + 2y\cos\theta\sin\theta\right)\mathbf{k}$$

and substituting some trigonometric identities, we obtain

$$q\mathbf{v}\bar{q} = x\mathbf{i} + (y\cos 2\theta - z\sin 2\theta)\,\mathbf{j} + (z\cos 2\theta + y\sin 2\theta)\,\mathbf{k} \quad (7.15)$$

and a pure rotation transform emerges with the rotation angle doubled:

$$\begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix}.$$

As the original quaternion product had the form

$$q\mathbf{v}\bar{q} = (\cos\theta + \sin\theta\mathbf{i})\,\mathbf{v}\,(\cos\theta - \sin\theta\mathbf{i})$$

and the subsequent rotation was about the x-axis, it seems reasonable to assume that a rotation about the y- and z-axes is achieved by

$$q\mathbf{v}\bar{q} = (\cos\theta + \sin\theta\mathbf{j})\,\mathbf{v}\,(\cos\theta - \sin\theta\mathbf{j})$$

and

$$q\mathbf{v}\bar{q} = (\cos\theta + \sin\theta\mathbf{k})\,\mathbf{v}\,(\cos\theta - \sin\theta\mathbf{k})$$

respectively, which is the case, and can be proved by repeating the above analysis with these quaternions.

For consistency, the vector **v** is represented as a pure quaternion where the real part is zero: $(0 + \mathbf{v})$. So, a rotation θ about an axis is achieved by

$$qv\bar{q} = \left[\cos\theta/2 + \sin\theta/2\,(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\right](0 + \mathbf{v})\left[\cos\theta/2 - \sin\theta/2\,(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\right] \quad (7.16)$$

where the axis of rotation is represented by a unit vector.

The above analysis has not been very rigorous; nevertheless, hopefully, it has shown that the algebra of quaternions is reasonably straightforward.

7.6 The complex conjugate of a quaternion

Now let's define the quaternion $\bar{q}$. The complex conjugate of an ordinary complex number is defined as

$$\bar{z} = a - b\mathbf{i}$$

where

$$z = a + b\mathbf{i}.$$

Similarly, given a quaternion

$$q = s + \mathbf{v}$$

its complex conjugate is defined as

$$\bar{q} = s - \mathbf{v}$$

which we innocently used in demonstrating the rotational properties of quaternions.

In Section 7.3.3 we proved that for two complex numbers

$$\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2.$$

Could this relationship hold for two quaternions? Let's see.

Given two quaternions

$$q_1 = s_1 + \mathbf{v}_1$$

and

$$q_2 = s_2 + \mathbf{v}_2$$

then

$$q_1 q_2 = (s_1 + \mathbf{v}_1)(s_2 + \mathbf{v}_2)$$

which must be expanded using the rules of the quaternion product:

$$q_1 q_2 = s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2 + s_1 \mathbf{v}_2 + s_2 \mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2.$$

Therefore,

$$\overline{q_1 q_2} = s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2 - s_1 \mathbf{v}_2 - s_2 \mathbf{v}_1 - \mathbf{v}_1 \times \mathbf{v}_2. \tag{7.17}$$

Now let's expand $\overline{q_1}\,\overline{q_2}$:

$$\overline{q_1}\,\overline{q_2} = (s_1 - \mathbf{v}_1)(s_2 - \mathbf{v}_2)$$

$$\overline{q_1}\,\overline{q_2} = s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2 - s_1 \mathbf{v}_2 - s_2 \mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2. \tag{7.18}$$

As we can see from Eqs. (7.17) and (7.18)

$$\overline{q_1 q_2} \neq \overline{q_1}\,\overline{q_2}.$$

But if we reverse q_1 and q_2 and expand $\overline{q_2}\,\overline{q_1}$, we find that

$$\overline{q_2}\,\overline{q_1} = (s_2 - \mathbf{v}_2)(s_1 - \mathbf{v}_1)$$
$$\overline{q_2}\,\overline{q_1} = s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2 - s_1 \mathbf{v}_2 - s_2 \mathbf{v}_1 - \mathbf{v}_1 \times \mathbf{v}_2$$

which proves that

$$\overline{q_1 q_2} = \overline{q_2}\,\overline{q_1}. \tag{7.19}$$

Now let's prove that

$$q\bar{q} = \bar{q}q.$$

Starting with

$$q\bar{q} = (s + \mathbf{v})(s - \mathbf{v})$$
$$q\bar{q} = s^2 - (\mathbf{v}) \cdot (-\mathbf{v}) - \mathbf{v} \times \mathbf{v} = s^2 + |\,\mathbf{v}\,|^2.$$

Similarly,

$$\bar{q}q = (s - \mathbf{v})(s + \mathbf{v})$$
$$\bar{q}q = s^2 - (-\mathbf{v}) \cdot \mathbf{v} - \mathbf{v} \times \mathbf{v} = s^2 + |\,\mathbf{v}\,|^2$$

Therefore,

$$q\bar{q} = \bar{q}q. \tag{7.20}$$

7.7 The norm of a quaternion

The *absolute value* or *magnitude* of a complex number

$$a + b\mathbf{i}$$

is defined as

$$|\,z\,| = \sqrt{a^2 + b^2}.$$

Similarly, the *norm* or *magnitude* of a quaternion

$$q = s + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

is defined as

$$|q| = \sqrt{s^2 + x^2 + y^2 + z^2}.$$

For example, given

$$q = 1 + 2\mathbf{i} - 3\mathbf{j} + 4\mathbf{k}$$

then

$$|q| = \sqrt{1^2 + 2^2 + (-3)^2 + 4^2} = \sqrt{30}.$$

If we represent the quaternion as

$$q = s + \mathbf{v}$$

then

$$|q| = \sqrt{s^2 + |\mathbf{v}|^2}$$

and from Eq. (7.20), we have

$$q\bar{q} = \bar{q}q = |q|^2. \tag{7.21}$$

Developing Eq. (7.21), we can state that

$$|q|^2 = q\bar{q}.$$

But as the single quaternion q can be a product of two quaternions

$$|q_1 q_2|^2 = (q_1 q_2)\,(\overline{q_1 q_2}).$$

Applying Eq. (7.19) gives

$$|q_1 q_2|^2 = q_1 q_2 \bar{q}_2 \bar{q}_1.$$

Applying Eq. (7.21) gives

$$|q_1 q_2|^2 = q_1 |q_2|^2 \bar{q}_1.$$

Rearranging gives

$$|q_1 q_2|^2 = q_1 \bar{q}_1 |q_2|^2.$$

Applying Eq. (7.21) gives

$$|q_1 q_2|^2 = |q_1|^2 |q_2|^2$$

Therefore,

$$|q_1 q_2| = |q_1|\,|q_2|$$

which informs us that the norm of a quaternion product equals the product of the individual norms.

As with unit vectors, there are unit quaternions, where $|q| = 1$, which are used extensively when rotating vectors.

7.8 Inverse of a quaternion

As we have seen in earlier chapters, vectors can be multiplied together, but division does not appear to have any meaning. Quaternions, on the other hand, can be divided, although the division is represented as a product using an inverse quaternion. For example, q^{-1} is used to represent $1/q$, which means that

$$qq^{-1} = 1. \tag{7.22}$$

In order to find q^{-1}, we multiply Eq. (7.22) by $\bar{q}$:

$$\bar{q}qq^{-1} = \bar{q}.$$

But from Eq. (7.21), we see that

$$\bar{q}q = |q|^2.$$

Therefore,

$$|q|^2 q^{-1} = \bar{q}$$

and

$$q^{-1} = \frac{\bar{q}}{|q|^2}.$$

If $|q| = 1$, then

$$q^{-1} = \bar{q}.$$

7.9 Rotating vectors using quaternions

Before discovering other features of quaternions, let's apply Eq. (7.16) to a practical example. To recap, the equation to rotate a vector is

$$qv\bar{q} = \left[\cos\theta/2 + \sin\theta/2\,(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\right](0 + \mathbf{v})\left[\cos\theta/2 - \sin\theta/2\,(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})\right] \tag{7.23}$$

where

θ is the angle of rotation,
$\mathbf{v}$ is the vector to be rotated, and
$x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ is the axis of rotation and in unit form.

The vector $\mathbf{v}$ is represented as a quaternion with a zero real component:

$$v = (0 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}).$$

Example 1

For the sake of simplicity, let's make the y-axis the axis of rotation and the angle of rotation 180°. Thus, if the position vector $\mathbf{v}$ is

$$\mathbf{v} = \mathbf{i} + 2\mathbf{j}$$

the rotating quaternions are

$$q = (0 + \mathbf{j})$$

and

$$\bar{q} = (0 - \mathbf{j}).$$

Note that the axis of rotation is represented by a unit vector.

Figure 7.6 illustrates this scenario.

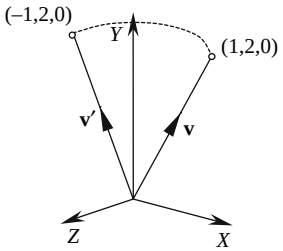


FIGURE 7.6.

Applying Eq. (7.23), we have

$$qv\bar{q} = (\cos 90° + \sin 90°\mathbf{j})\,(0 + \mathbf{i} + 2\mathbf{j})\,(\cos 90° - \sin 90°\mathbf{j})$$

$$qv\bar{q} = (0 + \mathbf{j})\,(0 + \mathbf{i} + 2\mathbf{j})\,(0 - \mathbf{j})$$

$$qv\bar{q} = (-2 - \mathbf{k})\,(0 - \mathbf{j})$$

$$qv\bar{q} = 0 + 2\mathbf{j} - \mathbf{i}.$$

Rearranging the terms gives

$$qv\bar{q} = 0 - \mathbf{i} + 2\mathbf{j}.$$

The vector part is

$$-\mathbf{i} + 2\mathbf{j}$$

which makes the rotated point $(-1, 2, 0)$, which is correct.

Example 2

Now let's rotate the same point about the axis $\frac{1}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})$:

$$qv\bar{q} = \left(\cos 90° + \frac{\sin 90°}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})\right)(0+\mathbf{i}+2\mathbf{j})\left(\cos 90° - \frac{\sin 90°}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})\right)$$

$$qv\bar{q} = \left(0 + \frac{1}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})\right)(0+\mathbf{i}+2\mathbf{j})\left(0 - \frac{1}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})\right)$$

$$qv\bar{q} = \left(-\frac{1}{\sqrt{2}} - \frac{3}{\sqrt{2}}\mathbf{k}\right)\left(0 - \frac{1}{\sqrt{2}}(-\mathbf{i}+\mathbf{j})\right)$$

$$qv\bar{q} = -\frac{1}{2}\mathbf{i} + \frac{1}{2}\mathbf{j} - \frac{3}{2}\mathbf{i} - \frac{3}{2}\mathbf{j}$$

$$qv\bar{q} = -2\mathbf{i} - \mathbf{j}$$

which is also correct, as seen from Fig. 7.7, which, for clarity, only shows the x- and y-axes.

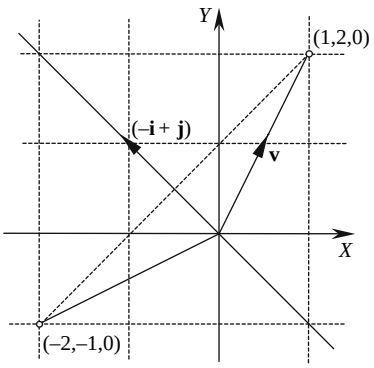


Figure 7.7.

7.10 Representing a quaternion as a matrix

Matrices play an important role in computer graphics, especially in representing transforms to manipulate objects and the virtual camera. Yaw, roll, and pitch matrices exist that rotate points about the three Cartesian axes and can be combined to create a single matrix, but, as mentioned above, can give rise to gimbal lock. Let us now derive a single matrix that will rotate a point an arbitrary axis without the problems of gimbal lock.

We begin by defining a quaternion q and its conjugate $\bar{q}$:

$$q = s + \mathbf{q} = s + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

$$\bar{q} = s - \mathbf{q} = s - x\mathbf{i} - y\mathbf{j} - z\mathbf{k}$$

where

$$s^2 + x^2 + y^2 + z^2 = 1$$

and the pure quaternion v, which contains the position vector $\mathbf{v}$:

$$v = 0 + \mathbf{v}.$$

Therefore,

$$qv\bar{q} = (s + \mathbf{q})(0 + \mathbf{v})(s - \mathbf{q}).$$

Multiplying the first two quaternions gives

$$qv\bar{q} = (-\mathbf{q}\cdot\mathbf{v} + s\mathbf{v} + \mathbf{q}\times\mathbf{v})(s - \mathbf{q}).$$

Multiplying these quaternions gives

$$qv\bar{q} = -\mathbf{q}\cdot\mathbf{v}s - (s\mathbf{v} + \mathbf{q}\times\mathbf{v})\cdot(-\mathbf{q}) + (-\mathbf{q}\cdot\mathbf{v})(-\mathbf{q}) + s(s\mathbf{v} + \mathbf{q}\times\mathbf{v}) + (s\mathbf{v} + \mathbf{q}\times\mathbf{v})\times(-\mathbf{q}).$$

Rearranging and simplifying terms gives

$$qv\bar{q} = -s\mathbf{q}\cdot\mathbf{v} + s\mathbf{v}\cdot\mathbf{q} + (-\mathbf{q}\cdot\mathbf{v})(-\mathbf{q}) + s^2\mathbf{v} + s\mathbf{q}\times\mathbf{v} + s\mathbf{v}\times(-\mathbf{q}) + \mathbf{q}\times\mathbf{v}\times(-\mathbf{q})$$
$$qv\bar{q} = (\mathbf{q}\cdot\mathbf{v})\mathbf{q} + s^2\mathbf{v} + s\mathbf{q}\times\mathbf{v} + s\mathbf{v}\times(-\mathbf{q}) + \mathbf{q}\times\mathbf{v}\times(-\mathbf{q}).$$

Using the triple vector product identity gives

$$\mathbf{q}\times\mathbf{v}\times(-\mathbf{q}) = -\mathbf{q}\cdot\mathbf{q}\mathbf{v} + (\mathbf{v}\cdot\mathbf{q})\mathbf{q}.$$

Therefore,

$$qv\bar{q} = (\mathbf{q}\cdot\mathbf{v})\mathbf{q} + s^2\mathbf{v} + s\mathbf{q}\times\mathbf{v} + s\mathbf{q}\times\mathbf{v} - \mathbf{q}\cdot\mathbf{q}\mathbf{v} + (\mathbf{v}\cdot\mathbf{q})\mathbf{q}$$
$$qv\bar{q} = \left(s^2 - |\mathbf{q}|^2\right)\mathbf{v} + 2(\mathbf{q}\cdot\mathbf{v})\mathbf{q} + 2s(\mathbf{q}\times\mathbf{v}). \tag{7.24}$$

The next task is to convert Eq. (7.24) into a matrix, which we will do in three steps.

Converting $\left(s^2 - |\mathbf{q}|^2\right)\mathbf{v}$:

$$|\mathbf{q}|^2 = x^2 + y^2 + z^2 = 1 - s^2.$$

Therefore,

$$s^2 - |\mathbf{q}|^2 = s^2 - \left(1 - s^2\right) = 2s^2 - 1$$

and

$$\left(s^2 - |\mathbf{q}|^2\right)\mathbf{v} = \left(2s^2 - 1\right)\mathbf{v} = \left(2s^2 - 1\right)x_v\mathbf{i} + \left(2s^2 - 1\right)y_v\mathbf{j} + \left(2s^2 - 1\right)z_v\mathbf{k}$$
$$\left(s^2 - |\mathbf{q}|^2\right)\mathbf{v} = \begin{bmatrix} 2s^2 - 1 & 0 & 0 \\ 0 & 2s^2 - 1 & 0 \\ 0 & 0 & 2s^2 - 1 \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}.$$

Converting $2(\mathbf{q}\cdot\mathbf{v})\,\mathbf{q}$:

$$2\mathbf{q}\cdot\mathbf{v} = 2(xx_v + yy_v + zz_v).$$

Therefore,

$$2(\mathbf{q}\cdot\mathbf{v})\,\mathbf{q} = 2(xx_v + yy_v + zz_v)(x\mathbf{i} + y\mathbf{j} + z\mathbf{k})$$
$$2(\mathbf{q}\cdot\mathbf{v})\,\mathbf{q} = 2\left(x^2 + xy + xz\right)x_v\mathbf{i} + 2\left(xy + y^2 + yz\right)y_v\mathbf{j} + 2\left(xz + yz + z^2\right)z_v\mathbf{k}$$

$$2(\mathbf{q}\cdot\mathbf{v})\,\mathbf{q} = \begin{bmatrix} 2x^2 & 2xy & 2xz \\ 2xy & 2y^2 & 2yz \\ 2xz & 2yz & 2z^2 \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}.$$

Converting $2s(\mathbf{q}\times\mathbf{v})$:

$$\mathbf{q}\times\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & z \\ x_v & y_v & z_v \end{vmatrix} = (yz_v - zy_v)\,\mathbf{i} + (zx_v - xz_v)\,\mathbf{j} - (xy_v - yx_v)\,\mathbf{k}$$
$$2s(\mathbf{q}\times\mathbf{v}) = 2s(yz_v - zy_v)\,\mathbf{i} + 2s(zx_v - xz_v)\,\mathbf{j} - 2s(xy_v - yx_v)\,\mathbf{k}$$
$$2s(\mathbf{q}\times\mathbf{v}) = \begin{bmatrix} 0 & -2sz & 2sy \\ 2sz & 0 & -2sx \\ -2sy & 2sx & 0 \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}.$$

Combining all three terms gives

$$qv\bar{q} = \begin{bmatrix} 2s^2 - 1 + 2x^2 & 2xy - 2sz & 2xz + 2sy \\ 2xy + 2sz & 2s^2 - 1 + 2y^2 & 2yz - 2sx \\ 2xz - 2sy & 2yz + 2sx & 2s^2 - 1 + 2z^2 \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}.$$

Simplifying leads to

$$qv\bar{q} = \begin{bmatrix} 2(s^2 + x^2) - 1 & 2(xy - sz) & 2(xz + sy) \\ 2(xy + sz) & 2(s^2 + y^2) - 1 & 2(yz - sx) \\ 2(xz - sy) & 2(yz + sx) & 2(s^2 + z^2) - 1 \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}$$

or

$$qv\bar{q} = \begin{bmatrix} 1 - 2(y^2 + z^2) & 2(xy - sz) & 2(xz + sy) \\ 2(xy + sz) & 1 - 2(x^2 + z^2) & 2(yz - sx) \\ 2(xz - sy) & 2(yz + sx) & 1 - 2(x^2 + y^2) \end{bmatrix}\begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}.$$

This is an important matrix as it is used to represent quaternion operations within computer graphics programs.

It will be interesting to see the matrix version of Example 2 above and show that it produces an identical result. To recap, the angle is 180°, the axis is

$$\frac{1}{\sqrt{2}}(-\mathbf{i} + \mathbf{j})$$

and the position vector to be rotated is

$$(1, 2, 0)$$

which makes

$$s = \cos 90^\circ = 0$$
$$x = -\frac{\sin 90^\circ}{\sqrt{2}} = -\frac{\sqrt{2}}{2}$$
$$y = \frac{\sin 90^\circ}{\sqrt{2}} = \frac{\sqrt{2}}{2}$$
$$z = 0$$

$$qv\bar{q} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \\ 0 \end{bmatrix} \tag{7.25}$$

which is the same result.

To summarize:

a position vector

$$\mathbf{v} = x_v\mathbf{i} + y_v\mathbf{j} + z_v\mathbf{k}$$

is rotated about an axis

$$\mathbf{q} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

either using

$$qv\bar{q} = \left(\cos\frac{\theta}{2} + \sin\frac{\theta}{2}\mathbf{q}\right)(0 + \mathbf{v})\left(\cos\frac{\theta}{2} - \sin\frac{\theta}{2}\mathbf{q}\right)$$

or

$$qv\bar{q} = \begin{bmatrix} 1 - 2(y^2 + z^2) & 2(xy - sz) & 2(xz + sy) \\ 2(xy + sz) & 1 - 2(x^2 + z^2) & 2(yz - sx) \\ 2(xz - sy) & 2(yz + sx) & 1 - 2(x^2 + y^2) \end{bmatrix} \begin{bmatrix} x_v \\ y_v \\ z_v \end{bmatrix}$$

where

$$s = \cos\frac{\theta}{2}$$

x, y, and z incorporate $\sin\frac{\theta}{2}$ and the axis components.

8 Vector Differentiation

8.1 Introduction

Since vectors were discovered, which is only about 150 years ago, a totally new world of mathematical analysis has emerged through their combination with differential and integral calculus. Differential vector analysis has provided new ways of describing tangential and normal vectors to curves and surfaces, as well as providing powerful analytical tools for studying scalar and vector fields. Integral vector analysis, on the other hand, has been applied to the study of vector fields used in fluid mechanics and heat transfer problems. Although many of these techniques are used in advanced aspects of computer graphics, in this introductory text we will explore only the fundamental concepts of differentiating vectors.

Readers who wish to explore this topic at an advanced level are recommended to study *Vector Analysis for Engineers and Scientists* by P. E. Lewis and J. P. Ward [Lewis, 1989]. That book was extremely useful in the preparation of this chapter.

8.2 The derivative of a vector

Differential calculus measures the rate of change of a function relative to one or more dependent variables. For example, the first differential of the polynomial

$$y = 2x^2 + 3x + 4$$

is

$$\frac{dy}{dx} = 4x + 3$$

which expresses how fast the function y is changing relative to the variable x for any value of x.

But what happens if the function is defined in terms of a vector equation such as

$$\mathbf{f}(t) = x(t)\,\mathbf{i} + y(t)\,\mathbf{j} + z(t)\,\mathbf{k}? \tag{8.1}$$

For example, Eq. (8.1) could describe a curve in R^3 such as a helix, and $\frac{df}{dt}$ would define the tangential slope of the curve at a point determined by t.

To define the first differential of a vector-based equation, consider the curve shown in Fig. 8.1, where two position vectors $\mathbf{f}(t)$ and $\mathbf{f}(t+\delta t)$ identify two points $P(t)$ and $Q(t+\delta t)$, respectively.

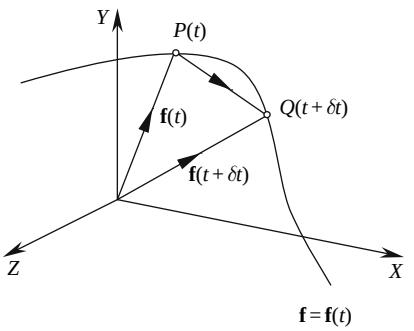


Figure 8.1.

From Fig. 8.1 we can state

$$\mathbf{f}(t) + \overrightarrow{PQ} = \mathbf{f}(t+\delta t).$$

Therefore,

$$\overrightarrow{PQ} = \mathbf{f}(t+\delta t) - \mathbf{f}(t).$$

If δt is small relative to the magnitude of $\overrightarrow{PQ}$, $\overrightarrow{PQ}$ can be replaced by $\delta\mathbf{f}$:

$$\delta\mathbf{f} = \mathbf{f}(t+\delta t) - \mathbf{f}(t). \tag{8.2}$$

Multiplying Eq. (8.2) by $\frac{1}{\delta t}$, we obtain

$$\frac{\delta\mathbf{f}}{\delta t} = \frac{\mathbf{f}(t+\delta t) - \mathbf{f}(t)}{\delta t}. \tag{8.3}$$

Expanding Eq. (8.3) into its components gives

$$\frac{\delta\mathbf{f}}{\delta t} = \frac{x(t+\delta t) - x(t)}{\delta t}\mathbf{i} + \frac{y(t+\delta t) - y(t)}{\delta t}\mathbf{j} + \frac{z(t+\delta t) - z(t)}{\delta t}\mathbf{k}.$$

In the limit as $\delta t \to 0$, we have

$$\lim_{\delta t \to 0} \frac{\delta \mathbf{f}}{\delta t} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k}.$$

Therefore,

$$\frac{d\mathbf{f}}{dt} = \lim_{\delta t \to 0} \frac{\delta \mathbf{f}}{\delta t} = \lim_{\delta t \to 0} \frac{\mathbf{f}(t + \delta t) - \mathbf{f}(t)}{\delta t}.$$

From Fig. 8.1 we can see that as Q approaches P, $\overrightarrow{QP}$ becomes tangential to the point P.

As an illustration, let's define a helix coiling upwards about the vertical y-axis:

$$\mathbf{f}(t) = \cos t\mathbf{i} + t\mathbf{j} + \sin t\mathbf{k}$$

and

$$\frac{d\mathbf{f}}{dt} = \frac{d}{dt}(\cos t)\mathbf{i} + \frac{d}{dt}(t)\mathbf{j} + \frac{d}{dt}(\sin t)\mathbf{k}$$
$$\frac{d\mathbf{f}}{dt} = (-\sin t)\mathbf{i} + \mathbf{j} + (\cos t)\mathbf{k}$$

which is the slope of the curve's tangent at the point $\mathbf{f}(t)$. For example, when $t = 2\pi$, we have

$$\frac{d\mathbf{f}}{dt} = \mathbf{j} + \mathbf{k}.$$

Without proof, the differentials of vector operations are

$$\frac{d}{dt}(\mathbf{p} + \mathbf{q}) = \frac{d\mathbf{p}}{dt} + \frac{d\mathbf{q}}{dt}$$

$$\frac{d}{dt}(\mathbf{p} \cdot \mathbf{q}) = \frac{d\mathbf{p}}{dt} \cdot \mathbf{q} + \mathbf{p} \cdot \frac{d\mathbf{q}}{dt}$$

$$\frac{d}{dt}(\mathbf{p} \times \mathbf{q}) = \frac{d\mathbf{p}}{dt} \times \mathbf{q} + \mathbf{p} \times \frac{d\mathbf{q}}{dt}$$

$$\frac{d}{dt}(\mathbf{p} \cdot \mathbf{q} \times \mathbf{r}) = \frac{d\mathbf{p}}{dt} \cdot \mathbf{q} \times \mathbf{r} + \mathbf{p} \cdot \frac{d\mathbf{q}}{dt} \times \mathbf{r} + \mathbf{p} \cdot \mathbf{q} \times \frac{d\mathbf{r}}{dt}$$

$$\frac{d}{dt}(\mathbf{p} \times (\mathbf{q} \times \mathbf{r})) = \frac{d\mathbf{p}}{dt} \times (\mathbf{q} \times \mathbf{r}) + \mathbf{p} \times \left(\frac{d\mathbf{q}}{dt} \times \mathbf{r}\right) + \mathbf{p} \times \left(\mathbf{q} \times \frac{d\mathbf{r}}{dt}\right).$$

8.3 The normal vector to a planar curve

Given a curve in R^2 of the form $y = f(x)$, we can construct a tangent vector as shown in Fig. 8.2, where

$$\mathbf{t} = \mathbf{i} + \frac{dy}{dx}\mathbf{j}. \tag{8.4}$$

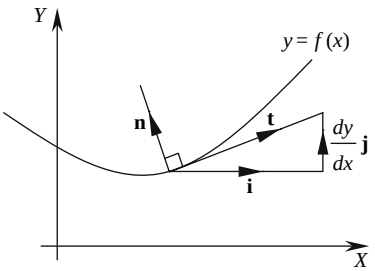


FIGURE 8.2.

However, in Section 2.10 we discovered that given a vector

$$\mathbf{v} = a\mathbf{i} + b\mathbf{j}$$

then

$$\mathbf{v}^{\perp} = -b\mathbf{i} + a\mathbf{j}.$$

Therefore, using Eq. (8.4), we have

$$\mathbf{n} = \mathbf{t}^{\perp} = -\frac{dy}{dx}\mathbf{i} + \mathbf{j}.$$

For example, given a function

$$y = 2x^2 + 3x - 4$$

$$\frac{dy}{dx} = 4x + 3.$$

Therefore,

$$\mathbf{n} = -(4x + 3)\,\mathbf{i} + \mathbf{j}$$

and when $x = 0$,

$$\mathbf{n} = -3\mathbf{i} + \mathbf{j}.$$

To develop a generic equation for line equations in R^2, we write the original curve equation as

$$y - f(x) = 0$$

or

$$F(x, y) = 0 \tag{8.5}$$

where the implicit function F is dependent upon two variables, but still describes a planar curve. If we make small changes to F using dx and dy, and remain on the curve, $F(x, y)$ must still equal zero, and $dF = 0$. Therefore,

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy = 0. \tag{8.6}$$

But Eq. (8.6) can be written as a vector scalar product:

$$dF = \left(\frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j}\right) \cdot (dx\mathbf{i} + dy\mathbf{j}) = 0.$$

Figure 8.3 illustrates why $dx\mathbf{i} + dy\mathbf{j}$ is the tangent vector to the curve F.

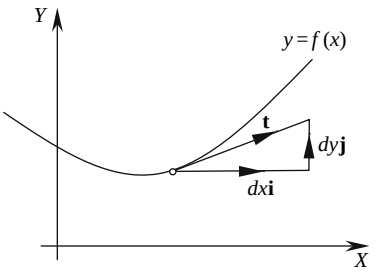


Figure 8.3.

If the product

$$\left(\frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j}\right) \cdot (dx\mathbf{i} + dy\mathbf{j}) = 0$$

then

$$\frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j}$$

must be the normal vector to the curve F:

$$\mathbf{n} = \frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j}.$$

So the curve F is defined as the scalar product of two perpendicular vectors: the tangent and normal vectors.

8.4 The normal vector to a surface

Surface normals are required in lighting calculations and are easily calculated from polygonal models. In this section we explore how normals are computed for continuous surfaces. We begin by transposing Eq. (8.5) into a 3D context to describe a surface:

$$F(x, y, z) = 0.$$

If we make small changes to F using dx, dy, and dz, and remain on the surface, $F(x, y, z)$ must still equal zero, and $dF = 0$. Therefore,

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial z}dz = 0. \tag{8.7}$$

But Eq. (8.7) can be written as vector scalar product:

$$dF = \left(\frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k}\right) \cdot (dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k}) = 0.$$

Similarly,

$$(dx\mathbf{i} + dy\mathbf{j} + dz\mathbf{k})$$

must be the tangent vector to the surface and, consequently,

$$\left(\frac{\partial F}{\partial x}\mathbf{i} + \frac{\partial F}{\partial y}\mathbf{j} + \frac{\partial F}{\partial z}\mathbf{k}\right)$$

must be the normal vector to the surface.

But say a surface is described in a parametric form such as a Bézier surface: then an alternative form is available. For example, let the position vector of a point on a surface be given by

$$\mathbf{r} = \mathbf{r}(u, v) = x(u, v)\,\mathbf{i} + y(u, v)\,\mathbf{j} + z(u, v)\,\mathbf{k}$$

as shown in Fig. 8.4.

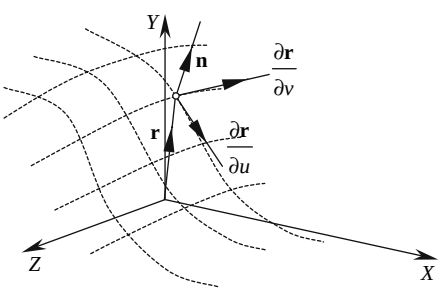


FIGURE 8.4.

Now, say we keep the parameter u constant ($u = c$):

$$\mathbf{r} = \mathbf{r}(c, v) = x(c, v)\,\mathbf{i} + y(c, v)\,\mathbf{j} + z(c, v)\,\mathbf{k}.$$

Then $\mathbf{r}$ is a function of v. $\partial\mathbf{r}/\partial v$ measures the rate of change of $\mathbf{r}$ relative to v and is represented by a tangential vector, as shown in Fig. 8.4. Similarly, by keeping the parameter v constant ($v = c$):

$$\mathbf{r} = \mathbf{r}(u, c) = x(u, c)\,\mathbf{i} + y(u, c)\,\mathbf{j} + z(u, c)\,\mathbf{k}$$

Then $\mathbf{r}$ is a function of u. $\partial\mathbf{r}/\partial u$ measures the rate of change of $\mathbf{r}$ relative to u and is represented by a tangential vector, as shown in Fig. 8.4.

Consequently, the two vectors $\partial\mathbf{r}/\partial u$ and $\partial\mathbf{r}/\partial v$ identify a tangential plane at the point $\mathbf{r}(u, v)$, which enables us to compute the normal vector $\mathbf{n}$:

$$\mathbf{n} = \frac{\partial\mathbf{r}}{\partial u} \times \frac{\partial\mathbf{r}}{\partial v}. \tag{8.8}$$

However, we must remember that $\mathbf{n}$ could also be defined as

$$\mathbf{n} = \frac{\partial\mathbf{r}}{\partial v} \times \frac{\partial\mathbf{r}}{\partial u}$$

as there is no right or wrong side to a surface patch. It is up to us to determine which side we wish represents the outside or inside.

Example 1

Consider, then, a quadratic Bézier patch defined as

$$\mathbf{r}(u, v) = \begin{bmatrix}(1-u)^2 & 2u(1-u) & u^2\end{bmatrix} \begin{bmatrix}\mathbf{p}_{00} & \mathbf{p}_{01} & \mathbf{p}_{02} \\ \mathbf{p}_{10} & \mathbf{p}_{11} & \mathbf{p}_{12} \\ \mathbf{p}_{20} & \mathbf{p}_{21} & \mathbf{p}_{22}\end{bmatrix} \begin{bmatrix}(1-v)^2 \\ 2v(1-v) \\ v^2\end{bmatrix}$$

where

$$\begin{aligned}\mathbf{p}_{00} &= (0,0,0) & \mathbf{p}_{01} &= (1,0,1) & \mathbf{p}_{02} &= (2,0,0) \\ \mathbf{p}_{10} &= (-1,1,1) & \mathbf{p}_{11} &= (1,1,3) & \mathbf{p}_{12} &= (3,1,1) \\ \mathbf{p}_{20} &= (0,2,0) & \mathbf{p}_{21} &= (1,3,1) & \mathbf{p}_{22} &= (2,2,0)\end{aligned}$$

$$\mathbf{r}(u, v) = \begin{bmatrix}\left(1-2u+u^2\right) & \left(2u-2u^2\right) & u^2\end{bmatrix} \begin{bmatrix}\mathbf{p}_{00} & \mathbf{p}_{01} & \mathbf{p}_{02} \\ \mathbf{p}_{10} & \mathbf{p}_{11} & \mathbf{p}_{12} \\ \mathbf{p}_{20} & \mathbf{p}_{21} & \mathbf{p}_{22}\end{bmatrix} \begin{bmatrix}(1-2v+v^2) \\ (2v-2v^2) \\ v^2\end{bmatrix}$$

$$x_r(u, v) = \begin{bmatrix}\left(1-2u+u^2\right) & \left(2u-2u^2\right) & u^2\end{bmatrix} \begin{bmatrix}0 & 1 & 2 \\ -1 & 1 & 3 \\ 0 & 1 & 2\end{bmatrix} \begin{bmatrix}(1-2v+v^2) \\ (2v-2v^2) \\ v^2\end{bmatrix}$$

$$x_r(u, v) = \begin{bmatrix}-\left(2u-2u^2\right)\end{bmatrix} \begin{bmatrix}(1-2u+u^2)+ \\ (2u-2u^2)+ \\ u^2\end{bmatrix} \begin{bmatrix}(2-4u+2u^2)+ \\ 6u-6u^2+ \\ 2u^2\end{bmatrix} \begin{bmatrix}(1-2v+v^2) \\ (2v-2v^2) \\ v^2\end{bmatrix}$$

$$x_r(u, v) = \left(2u^2-2u\right)(1)\left(-2u^2+2u+2\right) \begin{bmatrix}(1-2v+v^2) \\ (2v-2v^2) \\ v^2\end{bmatrix}$$

Therefore,

$$x_r(u,v) = \left(2u^2 - 2u\right)\left(1 - 2v + v^2\right) + \left(2v - 2v^2\right) + \left(-2u^2 + 2u + 2\right)\left(v^2\right)$$

and

$$y_r(u,v) = (2u)\left(1 - 2v + v^2\right) + \left(u^2 + 2u\right)\left(2v - 2v^2\right) + (2u)\left(v^2\right)$$
$$z_r(u,v) = \left(2u - 2u^2\right)\left(1 - 2v + v^2\right) + \left(1 + 4u - 4u^2\right)\left(2v - 2v^2\right) + \left(2u - 2u^2\right)\left(v^2\right).$$

Simplifying the terms gives

$$x_r(u,v) = 2u^2 - 4u^2v - 2u + 4uv + 2v$$
$$y_r(u,v) = 2u + 2u^2v - 2u^2v^2$$
$$z_r(u,v) = 2u + 2v + 4uv - 4uv^2 - 2u^2 - 2v^2 - 4u^2v + 4u^2v^2.$$

Expressing the preceding three equations as a vector equation gives

$$\begin{aligned}\mathbf{r} = {} & (2u^2 - 4u^2v - 2u + 4uv + 2v)\,\mathbf{i} \\ & + (2u + 2u^2v - 2u^2v^2)\,\mathbf{j} \\ & + (2u + 2v + 4uv - 4uv^2 - 2u^2 - 2v^2 - 4u^2v + 4u^2v^2)\,\mathbf{k}.\end{aligned}$$

Differentiating gives

$$\frac{\partial \mathbf{r}}{\partial u} = (4u - 8uv - 2 + 4v)\,\mathbf{i} + \left(2 + 4uv - 4uv^2\right)\mathbf{j} + \left(2 + 4v - 4v^2 - 4u - 8uv + 8uv^2\right)\mathbf{k}$$
$$\frac{\partial \mathbf{r}}{\partial v} = \left(-4u^2 + 4u + 2\right)\mathbf{i} + \left(2u^2 - 4u^2v\right)\mathbf{j} + \left(2 + 4u - 8uv - 4v - 4u^2 + 8u^2v\right)\mathbf{k}.$$

The differentials at $u = v = 0$ are

$$\frac{\partial \mathbf{r}}{\partial u} = -2\mathbf{i} + 2\mathbf{j} + 2\mathbf{k}$$
$$\frac{\partial \mathbf{r}}{\partial v} = 2\mathbf{i} + 2\mathbf{k}.$$

The surface normal is given by

$$\mathbf{n} = \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial r}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -2 & 2 & 2 \\ 2 & 0 & 2 \end{vmatrix}$$
$$\mathbf{n} = 4\mathbf{i} + 8\mathbf{j} - 4\mathbf{k}. \tag{8.9}$$

But how can we confirm that this is correct?

Well, a Bézier surface patch has the property that the surface is tangential to the mesh of corner control points, as shown in Fig. 8.5.

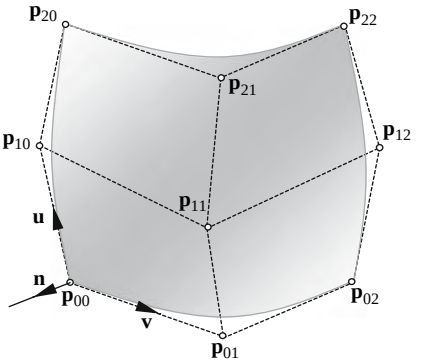


FIGURE 8.5.

The surface at $\mathbf{p}_{00}$ is tangential to the vectors $\mathbf{u}$ and $\mathbf{v}$. Therefore, the normal vector $\mathbf{n}$ is given by

$$\mathbf{n} = \mathbf{u} \times \mathbf{v}$$

where

$$\mathbf{u} = \mathbf{p}_{10} - \mathbf{p}_{00} = (-1 \quad 1 \quad 1) = -\mathbf{i} + \mathbf{j} + \mathbf{k}$$

and

$$\mathbf{v} = \mathbf{p}_{01} - \mathbf{p}_{00} = (1 \quad 0 \quad 1) = \mathbf{i}.$$

Therefore,

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -1 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix} = \mathbf{i} + 2\mathbf{j} - \mathbf{k}.$$

Although one would not expect the magnitude of this vector to equal that of Eq. (8.9), it must equal a scaled version. In this case the scaling factor is 4:

$$4(\mathbf{i} + 2\mathbf{j} - \mathbf{k}) = 4\mathbf{i} + 8\mathbf{j} - 4\mathbf{k}.$$

Example 2

Consider next the parametric equations for a sphere:

$$x = \sin u \cos v \quad y = \sin u \sin v \quad z = \cos u$$

or as a vector equation

$$\begin{gathered} \mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \\ \mathbf{r} = \sin u \cos v\mathbf{i} + y \sin u \sin v\mathbf{j} + \cos u\mathbf{k}. \end{gathered} \tag{8.10}$$

Therefore,

$$\frac{\partial \mathbf{r}}{\partial u} = \cos u \cos v\mathbf{i} + \cos u \sin v\mathbf{j} - \sin u\mathbf{k}$$

and

$$\frac{\partial \mathbf{r}}{\partial v} = -\sin u \sin v\mathbf{i} + \sin u \cos v\mathbf{j}.$$

Therefore, using Eq. (8.8) gives

$$\mathbf{n} = \frac{\partial \mathbf{r}}{\partial u} \times \frac{\partial \mathbf{r}}{\partial v}$$

$$\mathbf{n} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \cos u \cos v & \cos u \sin v & -\sin u \\ -\sin u \sin v & \sin u \cos v & 0 \end{vmatrix}$$

$$\mathbf{n} = \sin^2 u \cos v\mathbf{i} + \sin^2 u \sin v\mathbf{j} + \left(\sin u \cos u \cos^2 v + \cos u \sin u \sin^2 v\right)\mathbf{k}$$

$$\mathbf{n} = \sin^2 u \cos v\mathbf{i} + \sin^2 u \sin v\mathbf{j} + \sin u \cos u\mathbf{k}$$

$$\mathbf{n} = \sin u \left(\sin u \cos v\right)\mathbf{i} + \sin u \sin v\mathbf{j} + \cos u\mathbf{k}.$$

But from Eq. (8.10), we see that

$$\mathbf{r} = \sin u \cos v\mathbf{i} + y \sin u \sin v\mathbf{j} + \cos u\mathbf{k} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

in which case

$$\mathbf{n} = \sin u\mathbf{r}. \tag{8.11}$$

Equation (8.11) is just a scalar multiple of Eq. (8.10) and confirms that it represents a normal to the surface of a sphere.

Example 3

Let's find the unit tangent vector to any point on the curve

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$$

where

$$x = t^2 + 2 \quad y = 4t - 8 \quad z = 2t^2 - 4t.$$

Therefore,

$$\frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left[\left(t^2 + 2\right)\mathbf{i} + (4t - 8)\,\mathbf{j} + \left(2t^2 - 4t\right)\mathbf{k}\right]$$

$$\frac{d\mathbf{r}}{dt} = \frac{d}{dt}\left[2t\mathbf{i} + 4\mathbf{j} + (4t - 4)\,\mathbf{k}\right].$$

Thus, the tangential vector is

$$2t\mathbf{i} + 4\mathbf{j} + (4t - 4)\,\mathbf{k}.$$

Its magnitude is

$$\left\| \frac{d\mathbf{r}}{dt} \right\| = \sqrt{(2t)^2 + 16 + (4t - 4)^2}$$

and the unit tangential vector is

$$\hat{\mathbf{t}} = \frac{2t\mathbf{i} + 4\mathbf{j} + (4t - 4)\,\mathbf{k}}{\sqrt{(2t)^2 + 16 + (4t - 4)^2}}.$$

If we wanted to know the value of $\hat{\mathbf{t}}$ when $t = 1$, then we would find

$$\hat{\mathbf{t}} = \frac{2\mathbf{i} + 4\mathbf{j}}{\sqrt{4 + 16}} = \frac{2}{\sqrt{20}}\mathbf{i} + \frac{4}{\sqrt{20}}\mathbf{j}$$

$$\hat{\mathbf{t}} = \frac{1}{\sqrt{5}}\mathbf{i} + \frac{2}{\sqrt{5}}\mathbf{j}.$$

9 Projections

9.1 Introduction

Normally, one does not require vectors to derive the basic perspective transform — simple ratios are sufficient, but if the projection plane is rotated or oblique to the view direction, vector analysis comes to the rescue. So let's investigate how vectors can resolve some aspects of projections.

9.2 Perspective transform

Figure 9.1 assumes that an observer is located at the origin O and directed along the z-axis. Consequently, any point $P(x_P, y_P, z_P)$ in R^3 can be projected onto a projection plane revealing a point Q in R^2. The projection plane is orthogonal to the view direction, and the screen axes are parallel with the camera axes.

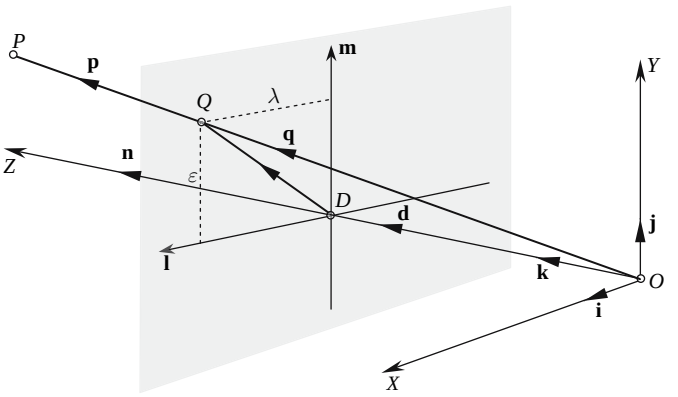


Figure 9.1.

We begin by declaring the following conditions:

$$\mathbf{l} = \mathbf{i} \qquad \mathbf{m} = \mathbf{j} \qquad \mathbf{n} = \mathbf{k}$$

$$\mathbf{d} = \overrightarrow{OD} = d\mathbf{k} \qquad \mathbf{q} = \overrightarrow{OQ} \qquad \mathbf{p} = \overrightarrow{OP}$$

and $Q(\lambda, \varepsilon)$ is the point on the projection plane intersected by $\mathbf{p}$. Our task is to find the scalars λ and ε.

From Fig. 9.1 we see that

$$\mathbf{q} = \mathbf{d} + \overrightarrow{DQ}$$

but $\overrightarrow{DQ}$ has local components:

$$\overrightarrow{DQ} = \lambda\mathbf{l} + \varepsilon\mathbf{m} = \lambda\mathbf{i} + \varepsilon\mathbf{j}.$$

Therefore,

$$\mathbf{q} = \mathbf{d} + \lambda\mathbf{i} + \varepsilon\mathbf{j}$$

or

$$\lambda\mathbf{i} + \varepsilon\mathbf{j} = \mathbf{q} - \mathbf{d}. \tag{9.1}$$

We now define $\mathbf{q}$ in terms of $\mathbf{p}$, so we let

$$\mathbf{q} = \kappa\mathbf{p}$$

where κ is a scalar that requires defining.

As $\mathbf{q}$ and $\kappa\mathbf{p}$ have the same projection on $\mathbf{k}$, we find that

$$\mathbf{k} \cdot \mathbf{q} = \kappa\mathbf{k} \cdot \mathbf{p}$$

and

$$\kappa = \frac{\mathbf{k} \cdot \mathbf{q}}{\mathbf{k} \cdot \mathbf{p}} = \frac{d}{z_p}.$$

Substituting κ in Eq. (9.1) gives

$$\lambda\mathbf{i} + \varepsilon\mathbf{j} = \frac{d}{z_p}(x_p\mathbf{i} + y_p\mathbf{j} + z_p\mathbf{k}) - d\mathbf{k}.$$

Equating the $\mathbf{i}$ and $\mathbf{j}$ components gives

$$\lambda = d\frac{x_p}{z_p}, \qquad \varepsilon = d\frac{y_p}{z_p}$$

which are readily recognised as the perspective projection plane coordinates.

9.2.1 Horizontally oblique projection plane

Now let's consider the case where the projection plane is rotated θ about the vertical $\mathbf{m}$-axis, as shown in Fig. 9.2.

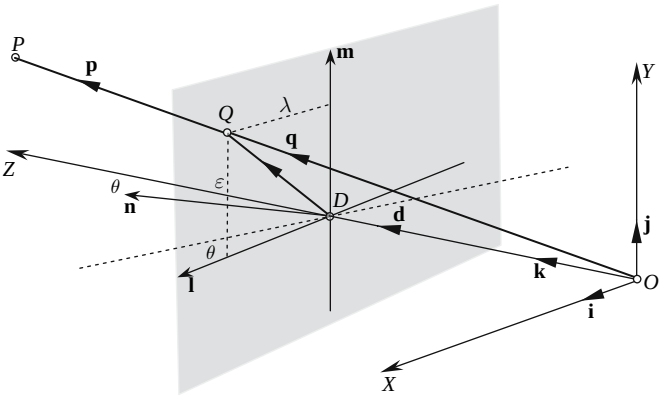


FIGURE 9.2.

We begin by declaring the following conditions:

$$\mathbf{i}\cdot\mathbf{l}=\cos\theta \qquad \mathbf{m}=\mathbf{j} \qquad \mathbf{n}\cdot\mathbf{k}=\cos\theta$$

$$\mathbf{d}=\overrightarrow{OD}=d\mathbf{k} \qquad \mathbf{q}=\overrightarrow{OQ} \qquad \mathbf{p}=\overrightarrow{OP}$$

$$\mathbf{n}\cdot\mathbf{d}=d\cos\theta.$$

The relationship between the rotated axial system **lmn** and **ijk** can be represented as

$$\begin{bmatrix}\mathbf{l}\\ \mathbf{m}\\ \mathbf{n}\end{bmatrix}=\begin{bmatrix}\cos\theta & 0 & -\sin\theta\\ 0 & 1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix}\cdot\begin{bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{bmatrix} \tag{9.2}$$

and represents the transform relating coordinates in **ijk** to **lmn**. $Q(\lambda,\ \varepsilon)$ is the point on the projection plane intersected by **p**, and once more, our task is to find the scalars λ and ε.

From Fig. 9.2 we see that

$$\mathbf{q}=\mathbf{d}+\overrightarrow{DQ}$$

but

$$\overrightarrow{DQ}=\lambda\mathbf{l}+\varepsilon\mathbf{m}.$$

Therefore,

$$\mathbf{q}=\mathbf{d}+\lambda\mathbf{l}+\varepsilon\mathbf{m}$$

or

$$\lambda\mathbf{l}+\varepsilon\mathbf{m}=\mathbf{q}-\mathbf{d}. \tag{9.3}$$

We now define **q** in terms of **p**, so we let

$$\mathbf{q}=\kappa\mathbf{p}$$

where κ is a scalar that requires defining.

As $\mathbf{q}$ and $\kappa\mathbf{p}$ have the same projection on $\mathbf{n}$, then

$$\mathbf{n}\cdot\mathbf{q} = \kappa\mathbf{n}\cdot\mathbf{p}$$

and

$$\kappa = \frac{\mathbf{n}\cdot\mathbf{d}}{\mathbf{n}\cdot\mathbf{p}} = \frac{d\mathbf{n}\cdot\mathbf{k}}{\mathbf{n}\cdot\mathbf{p}}.$$

We now define $\mathbf{n}$ in terms of $\mathbf{ijk}$. Therefore, from Eq. (9.2), we find that

$$\mathbf{n} = \sin\theta\mathbf{i} + \cos\theta\mathbf{k}$$

and

$$\kappa = \frac{d\cos\theta}{x_P\sin\theta + z_P\cos\theta}.$$

Returning to Eq. (9.3), we substitute $\mathbf{l}$ and $\mathbf{m}$ in terms of $\mathbf{ijk}$.

From Eq. (9.2), we have

$$\mathbf{l} = \cos\theta\mathbf{i} - \sin\theta\mathbf{k}$$

$$\mathbf{m} = \mathbf{j}$$

and Eq. (9.3) becomes

$$\lambda(\cos\theta\mathbf{i} - \sin\theta\mathbf{k}) + \varepsilon\mathbf{j} = \kappa(x_P\mathbf{i} + y_P\mathbf{j} + z_P\mathbf{k}) - d\mathbf{k}. \tag{9.4}$$

We now isolate the $\mathbf{i}$ components of Eq. (9.4) to reveal λ:

$$\lambda\cos\theta = \kappa x_P$$

and

$$\lambda = \frac{\kappa x_P}{\cos\theta}.$$

Finally, substituting κ gives

$$\lambda = \frac{d\cos\theta}{x_P\sin\theta + z_P\cos\theta}\left(\frac{x_P}{\cos\theta}\right)$$

$$\lambda = \frac{d\,x_P}{x_P\sin\theta + z_P\cos\theta}.$$

We now isolate the $\mathbf{j}$ components of Eq. (9.4) to reveal ε:

$$\varepsilon = \kappa y_P$$

and substituting κ gives

$$\varepsilon = \frac{d\,y_P\cos\theta}{x_P\sin\theta + z_P\cos\theta}.$$

Thus, we have found $Q(\lambda,\ \varepsilon)$. Note that when $\theta = 0$, λ and ε equal the values for the orthogonal projection plane.

9.2.2 Vertically oblique projection plane

Now let's consider the case where the projection plane is rotated θ about the horizontal **l**-axis, as shown in Fig. 9.3.

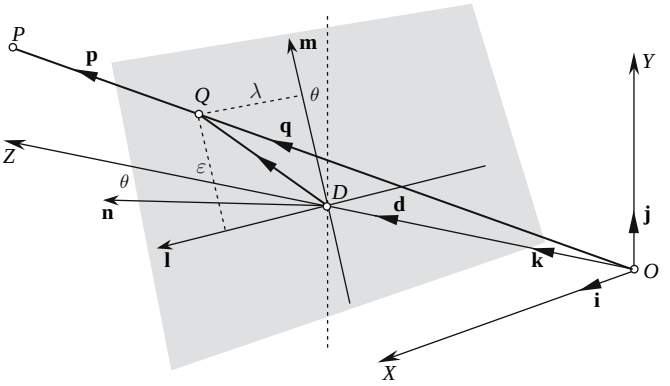


Figure 9.3.

We begin by declaring the following conditions:

$$\mathbf{j}\cdot\mathbf{m}=\cos\theta \qquad \mathbf{l}=\mathbf{i} \qquad \mathbf{m}\cdot\mathbf{i}=\cos\theta$$
$$\mathbf{d}=\overrightarrow{OD}=d\mathbf{k} \qquad \mathbf{q}=\overrightarrow{OQ} \qquad \mathbf{p}=\overrightarrow{OP}$$
$$\mathbf{n}\cdot\mathbf{d}=d\cos\theta.$$

The relationship between the rotated axial system **lmn** and **ijk** can be represented as

$$\begin{bmatrix}\mathbf{l}\\ \mathbf{m}\\ \mathbf{n}\end{bmatrix}=\begin{bmatrix}1 & 0 & 0\\ 0 & \cos\theta & \sin\theta\\ 0 & -\sin\theta & \cos\theta\end{bmatrix}\cdot\begin{bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{bmatrix} \tag{9.5}$$

and represents the transform relating coordinates in **ijk** to **lmn**. $Q(\lambda,\ \varepsilon)$ is the point on the projection plane intersected by **p**, and once more, our task is to find the scalars λ and ε.

From Fig. 9.3, we have

$$\mathbf{q}=\mathbf{d}+\overrightarrow{DQ}$$

but

$$\overrightarrow{DQ}=\lambda\mathbf{l}+\varepsilon\mathbf{m}$$

and

$$\mathbf{q}=\mathbf{d}+\lambda\mathbf{l}+\varepsilon\mathbf{m}=\lambda\mathbf{i}+\varepsilon\mathbf{m}$$

or

$$\lambda \mathbf{i} + \varepsilon \mathbf{m} = \mathbf{q} - \mathbf{d}. \tag{9.6}$$

We again define $\mathbf{q}$ in terms of $\mathbf{p}$, so we let

$$\mathbf{q} = \kappa \mathbf{p}$$

where κ is a scalar that requires defining.

As $\mathbf{q}$ and $\kappa\mathbf{p}$ have the same projection on $\mathbf{n}$, we have

$$\mathbf{n} \cdot \mathbf{q} = \kappa \mathbf{n} \cdot \mathbf{p}$$

and

$$\kappa = \frac{\mathbf{n} \cdot \mathbf{d}}{\mathbf{n} \cdot \mathbf{p}} = \frac{d\mathbf{n} \cdot \mathbf{k}}{\mathbf{n} \cdot \mathbf{p}}.$$

We now define $\mathbf{n}$ in terms of $\mathbf{ijk}$. Therefore, from Eq. (9.5), we have

$$\mathbf{n} = -\sin\theta \mathbf{j} + \cos\theta \mathbf{k}$$

and

$$\kappa = \frac{d\cos\theta}{-y_P \sin\theta + z_P \cos\theta}.$$

Returning to Eq. (9.6), we substitute $\mathbf{m}$ in terms of $\mathbf{ijk}$.

From Eq. (9.5), we have

$$\mathbf{m} = \cos\theta \mathbf{j} + \sin\theta \mathbf{k}$$

and Eq. (9.6) becomes

$$\lambda \mathbf{i} + \varepsilon (\cos\theta \mathbf{j} + \sin\theta \mathbf{k}) = \kappa (x_P \mathbf{i} + y_P \mathbf{j} + z_P \mathbf{k}) - d\mathbf{k}. \tag{9.7}$$

We now isolate the $\mathbf{i}$ components of Eq. (9.7) to reveal λ:

$$\lambda = \kappa x_P.$$

Finally, substituting κ gives

$$\lambda = \frac{d x_P \cos\theta}{-y_P \sin\theta + z_P \cos\theta}.$$

We isolate the $\mathbf{j}$ components of Eq. (9.7) to reveal ε:

$$\varepsilon = \frac{\kappa y_P}{\cos\theta}$$

and substituting κ gives

$$\varepsilon = \frac{d\cos\theta}{-y_P \sin\theta + z_P \cos\theta} \left(\frac{y_P}{\cos\theta}\right)$$

$$\varepsilon = \frac{d y_P}{-y_P \sin\theta + z_P \cos\theta}.$$

Note that when $\theta = 0$, λ and ε equal the values for the orthogonal projection plane.

Having examined two orientations of the projection plane, let's complete this analysis with an arbitrary orientation of the plane.

9.2.3 Arbitrary orientation of the projection plane

The above technique of using a rotation matrix to relate the **lmn**-axes with the **ijk**-axes provides the solution to this problem. This time the projection plane is subjected to yaw, pitch, and roll rotations, as shown in Fig. 9.4.

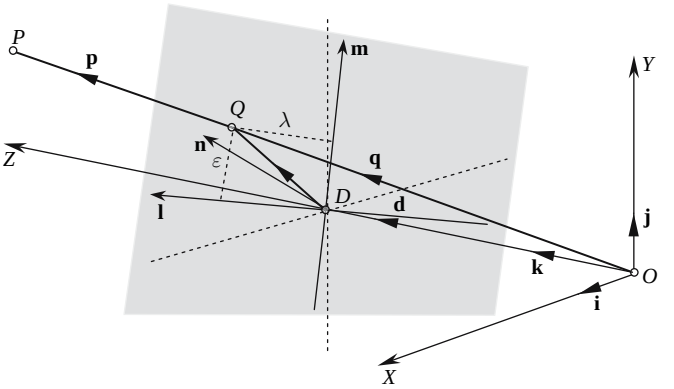


FIGURE 9.4.

The transform relating coordinates in **ijk** with **lmn** is given by

$$\begin{bmatrix} \mathbf{l} \\ \mathbf{m} \\ \mathbf{n} \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{bmatrix}. \tag{9.8}$$

We begin by declaring the following conditions:

$$\mathbf{d} = \overrightarrow{OD} = d\mathbf{k} \qquad \mathbf{q} = \overrightarrow{OQ} \qquad \mathbf{p} = \overrightarrow{OP}$$

$Q(\lambda, \varepsilon)$ is the point on the projection plane intersected by **p**, and once more, our task is to find the scalars λ and ε.

From Fig. 9.4, we see that

$$\mathbf{q} = \mathbf{d} + \overrightarrow{DQ}$$

but

$$\overrightarrow{DQ} = \lambda\mathbf{l} + \varepsilon\mathbf{m}.$$

Therefore,

$$\mathbf{q} = \mathbf{d} + \lambda\mathbf{l} + \varepsilon\mathbf{m}$$

or

$$\lambda\mathbf{l} + \varepsilon\mathbf{m} = \mathbf{q} - \mathbf{d}. \tag{9.9}$$

We again define $\mathbf{q}$ in terms of $\mathbf{p}$, so we let

$$\mathbf{q} = \kappa\mathbf{p}$$

where κ is a scalar that requires defining.

As $\mathbf{q}$ and $\kappa\mathbf{p}$ have the same projection on $\mathbf{n}$, we have

$$\mathbf{n}\cdot\mathbf{q} = \kappa\mathbf{n}\cdot\mathbf{p}$$

and

$$\kappa = \frac{\mathbf{n}\cdot\mathbf{d}}{\mathbf{n}\cdot\mathbf{p}} = \frac{d\mathbf{n}\cdot\mathbf{k}}{\mathbf{n}\cdot\mathbf{p}}.$$

We now define $\mathbf{n}$ in terms of $\mathbf{ijk}$. Therefore, from Eq. (9.8), we find

$$\mathbf{n} = t_{31}\mathbf{i} + t_{32}\mathbf{j} + t_{33}\mathbf{k}$$

and

$$\kappa = \frac{d\,t_{33}}{x_P t_{31} + y_P t_{32} + z_P t_{33}}$$

and Eq. (9.9) becomes

$$\lambda\mathbf{l} + \varepsilon\mathbf{m} = \kappa\mathbf{p} - d\mathbf{k}. \tag{9.10}$$

We can isolate λ by multiplying Eq. (9.10) throughout by $\mathbf{l}$

$$\lambda\mathbf{l}\cdot\mathbf{l} + \varepsilon\mathbf{l}\cdot\mathbf{m} = \mathbf{l}\cdot(\kappa\mathbf{p} - d\mathbf{k})$$
$$\lambda = \kappa\mathbf{l}\cdot\mathbf{p} - d\mathbf{l}\cdot\mathbf{k}.$$

But from Eq. (9.8) , we have

$$\mathbf{l} = t_{11}\mathbf{i} + t_{12}\mathbf{j} + t_{13}\mathbf{k}.$$

Therefore,

$$\lambda = \kappa\,(t_{11}\mathbf{i} + t_{12}\mathbf{j} + t_{13}\mathbf{k})\cdot\mathbf{p} - d\,(t_{11}\mathbf{i} + t_{12}\mathbf{j} + t_{13}\mathbf{k})\cdot\mathbf{k}$$
$$\lambda = \kappa\,(t_{11}\mathbf{i} + t_{12}\mathbf{j} + t_{13}\mathbf{k})\cdot(x_P\mathbf{i} + y_P\mathbf{j} + z_P\mathbf{k}) - dt_{13}$$
$$\lambda = \kappa\,(x_P t_{11} + y_P t_{12} + z_P t_{13}) - dt_{13}.$$

Finally, substituting κ gives

$$\lambda = \frac{dt_{33}}{x_P t_{31} + y_P t_{32} + z_P t_{33}}\,(x_P t_{11} + y_P t_{12} + z_P t_{13}) - dt_{13}$$
$$\lambda = dt_{33}\frac{x_P t_{11} + y_P t_{12} + z_P t_{13}}{x_P t_{31} + y_P t_{32} + z_P t_{33}} - dt_{13}.$$

We isolate ε by multiplying Eq. (9.10) throughout by $\mathbf{m}$

$$\lambda \mathbf{m}\cdot\mathbf{l}+\varepsilon \mathbf{m}\cdot\mathbf{m}=\mathbf{m}\cdot(\kappa\mathbf{p}-d\mathbf{k})=\varepsilon$$

$$\varepsilon=\mathbf{m}\cdot(\kappa\mathbf{p}-d\mathbf{k})\,. \tag{9.11}$$

But from Eq. (9.8), we find that

$$\mathbf{m}=t_{21}\mathbf{i}+t_{22}\mathbf{j}+t_{23}\mathbf{k}$$

Therefore,

$$\varepsilon=(t_{21}\mathbf{i}+t_{22}\mathbf{j}+t_{23}\mathbf{k})\cdot(\kappa\mathbf{p}-d\mathbf{k})$$
$$\varepsilon=\kappa\,(t_{21}\mathbf{i}+t_{22}\mathbf{j}+t_{23}\mathbf{k})\cdot(x_P\mathbf{i}+y_P\mathbf{j}+z_P\mathbf{k})-dt_{23}$$

and substituting κ gives

$$\varepsilon=\frac{d\,t_{33}}{x_Pt_{31}+y_Pt_{32}+z_Pt_{33}}\,(t_{21}\mathbf{i}+t_{22}\mathbf{j}+t_{23}\mathbf{k})\cdot(x_P\mathbf{i}+y_P\mathbf{j}+z_P\mathbf{k})-dt_{23}$$
$$\varepsilon=d\,t_{33}\frac{x_Pt_{21}+y_Pt_{22}+z_Pt_{23}}{x_Pt_{31}+y_Pt_{32}+z_Pt_{33}}-dt_{23}.$$

The reader may wish to confirm that the values of λ and ε reproduce the previously computed values for different transforms relating **lmn** with **ijk**.

Note that the transform relating **lmn** with **ijk** is derived by concatenating the yaw, roll, and pitch transforms associated with changing coordinates between two frames of reference. For instance, the individual transforms are given by

$$\begin{bmatrix}\cos roll & \sin roll & 0\\ -\sin roll & \cos roll & 0\\ 0 & 0 & 1\end{bmatrix}\quad\begin{bmatrix}1 & 0 & 0\\ 0 & \cos pitch & \sin pitch\\ 0 & -\sin pitch & \cos pitch\end{bmatrix}\quad\begin{bmatrix}\cos yaw & 0 & -\sin yaw\\ 0 & 1 & 0\\ \sin yaw & & \cos yaw\end{bmatrix}$$

and are concatenated as follows:

$$\begin{bmatrix}\mathbf{l}\\ \mathbf{m}\\ \mathbf{n}\end{bmatrix}=[roll]\;[pitch]\;[yaw]\begin{bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{bmatrix}$$
$$\begin{bmatrix}\mathbf{l}\\ \mathbf{m}\\ \mathbf{n}\end{bmatrix}=\begin{bmatrix}t_{11} & t_{12} & t_{13}\\ t_{21} & t_{22} & t_{23}\\ t_{31} & t_{32} & t_{33}\end{bmatrix}\cdot\begin{bmatrix}\mathbf{i}\\ \mathbf{j}\\ \mathbf{k}\end{bmatrix}$$

where

$$t_{11}=\cos yaw\;\cos roll+\sin yaw\;\sin pitch\;\sin roll$$
$$t_{12}=\cos pitch\;\sin roll$$
$$t_{13}=-\sin yaw\;\cos roll+\cos yaw\;\sin pitch\;\sin roll$$

$$t_{21} = -\cos yaw\ \sin roll + \sin yaw\ \sin pitch\ \cos roll$$

$$t_{22} = \cos pitch\ \cos roll$$

$$t_{23} = \sin yaw\ \sin roll + \cos yaw\ \sin pitch\ \cos roll$$

$$t_{31} = \sin yaw\ \cos pitch$$

$$t_{32} = -\sin pitch$$

$$t_{33} = \cos yaw\ \cos pitch$$

Figure 9.5 shows four views of a cube using the above equations: (a) shows the normal perspective projection; (b) shows the projection plane with a horizontal yaw of 10°; (c) shows the projection plane with a pitch of 10°; and (d) shows the projection plane subjected to a roll, pitch and yaw of 10°. Note, also, that the x coordinate has been reversed, to compensate for the reversal that occurs when the image is displayed with the x-axis pointing to the right.

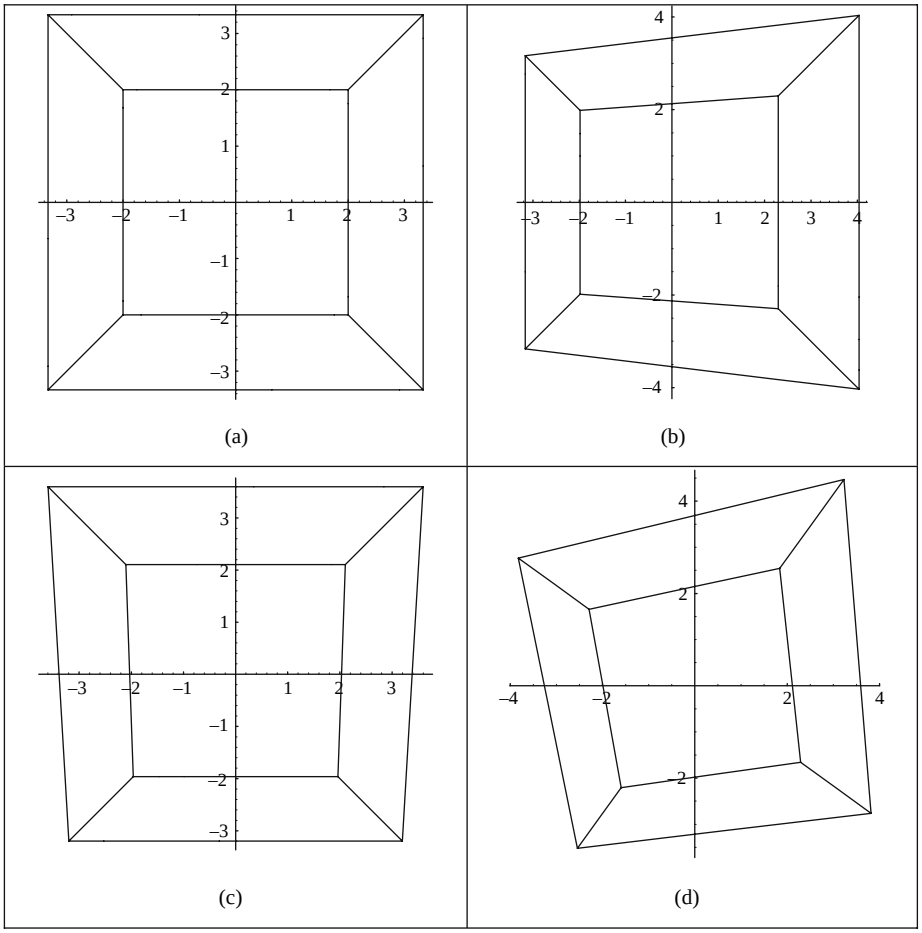


Figure 9.5.

9.3 Pseudo fish-eye projection

A pseudo fish-eye projection is readily achieved by projecting onto a spherical projection surface. The geometry for this is shown in Fig. 9.6, where the projection radius of curvature is given by r and the centre of curvature is located at the origin.

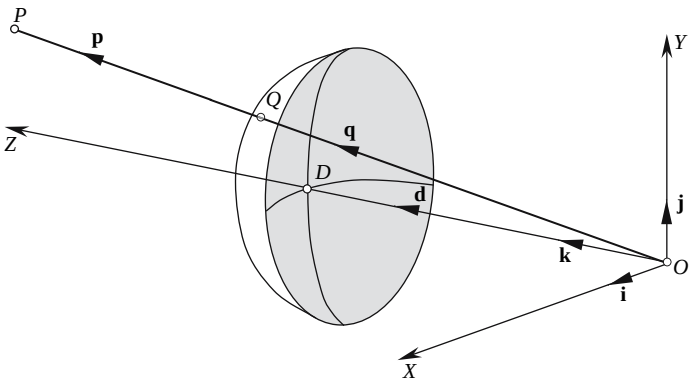


Figure 9.6.

We begin by declaring the following conditions:

$$\mathbf{d} = \overrightarrow{OD} = d\mathbf{k} \qquad \mathbf{q} = \overrightarrow{OQ} = r \qquad \mathbf{p} = \overrightarrow{OP}.$$

We define $\mathbf{q}$ in terms of $\mathbf{p}$, so we let

$$\mathbf{q} = \kappa\mathbf{p}. \tag{9.12}$$

But it is obvious that $\| \mathbf{q} \| = r$ for all points. Therefore,

$$\| \mathbf{q} \| = \kappa \| \mathbf{p} \|$$

and

$$\kappa = \frac{\| \mathbf{q} \|}{\| \mathbf{p} \|} = \frac{r}{\| \mathbf{p} \|}.$$

Substituting κ in Eq. (9.12) gives

$$\mathbf{q} = \frac{r}{\| \mathbf{p} \|}\mathbf{p}.$$

Therefore,

$$x_Q = \frac{rx_P}{\sqrt{x_P^2 + y_P^2 + z_P^2}} \qquad y_Q = \frac{ry_P}{\sqrt{x_P^2 + y_P^2 + z_P^2}}.$$

where r is the radius of curvature of the lens.

The effect of these equations is shown in Fig. 9.7, where a cube is viewed through a projection screen with a radius of curvature of 10. Care must be taken when using this projection because if only the end points of each edge are processed, then straight edges transform into straight edges. However, in Fig. 9.7 many points are taken from each edge to create the curved lines.

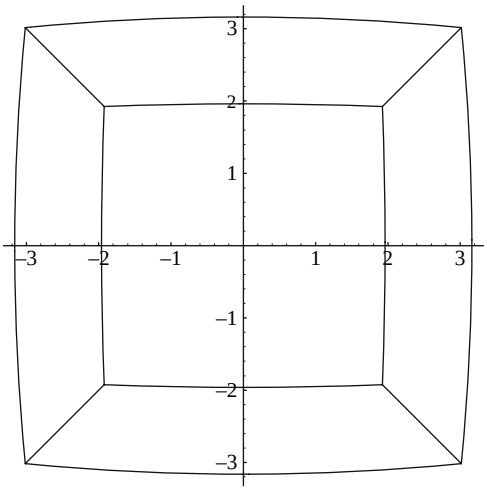


Figure 9.7.

10 Rendering

10.1 Introduction

The rendering phase generates images from geometric description of 3D models using a collection of virtual light sources. This requires knowledge of the location and type of the light sources, a description of the illumination model, and knowledge of how light intensities are mapped onto the geometry.

Vectors play an important role in this phase as they reveal the orientation of surfaces as well as the direction of light sources and assist in the analysis of how light interacts with objects. So let's begin with light sources.

10.2 Light sources

Light sources are often classified as point, directional, or spot. A point light source has spatial position and radiates light uniformly in all directions; a directional light source is located at infinity and gives rise to a parallel beam of light rays; a spotlight behaves like a torch, which has position, direction, and an angle associated with its conical beam.

As light sources involve parameters such as direction and magnitude, vectors are the perfect mathematical tool to describe and analyze their behavior. Although a vector is used to encode direction, individual scalars represent intensity and color; however, there is no reason why these color values could not also be treated as a vector. To simulate the influence a point light source has on a surface, we only need to define a vector between the position of the light source and some point on the illuminated surface. A vector $\mathbf{L}$ is used to represent a beam of photons (light) leaving the light source L and arriving at the point P, as shown in Fig. 10.1.
It is convenient to arrange for the vector $\mathbf{L}$ to point towards the light source, which is given by

$$\mathbf{L} = \begin{bmatrix} L_X - P_X \\ L_Y - P_Y \\ L_Z - P_Z \end{bmatrix}$$

where L and P locate the light source and illuminated point, respectively.

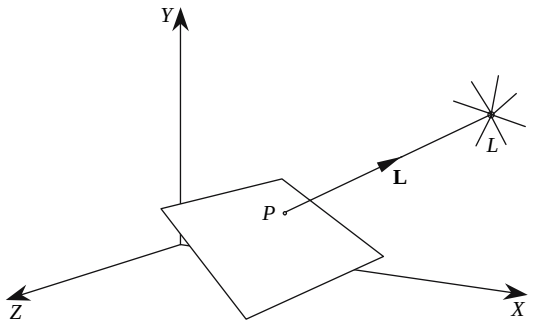


FIGURE 10.1.

In physics, two models are used to describe light: corpuscular and waves. The corpuscular model assumes that light is made up of small packets of electromagnetic energy called photons, which move at the speed of light. This is a convenient model to use when trying to understand the intensity of an illuminated surface. For example, when we observe an illuminated surface, its intensity is determined by the number of photons we perceive per unit surface area. The more photons detected, the more intense the surface appears. Imagine, then, a matte surface illuminated by a point light source, as shown in Fig. 10.2. Photons striking the surface are reflected back equally in all directions due to the random nature of the surface geometry.

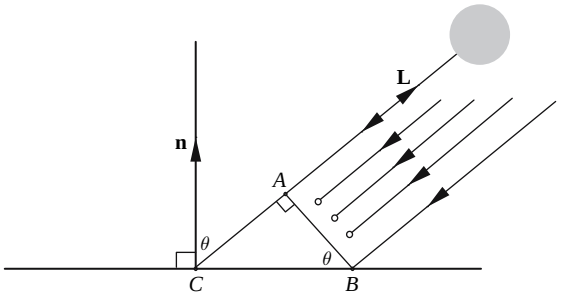


FIGURE 10.2.

If the line AB represents the diameter of a 3D circle, and the line CB represents the diameter of a 3D circle on the matte surface, then the number of photons n per unit area striking CB is proportional to

$$n \propto \frac{AB}{CB}.$$

But from Fig. 10.2, we have

$$\frac{AB}{CB} = \cos\theta$$

Therefore,

$$n \propto \cos\theta.$$

But $\cos\theta$ can be computed from the dot product using **n** and **L**:

$$\cos\theta = \frac{\mathbf{n}\cdot\mathbf{L}}{\|\mathbf{n}\|\,\|\mathbf{L}\|}$$

and if **n** and **L** are unit vectors, then

$$\cos\theta = \mathbf{n}\cdot\mathbf{L}$$

which allows us to state that the surface intensity I of a diffuse surface obeys the following law:

$$I = I_i\mathbf{n}\cdot\mathbf{L}$$

where

I_i is the incident light intensity,
n is the unit surface normal, and
L is the unit vector from the surface to the light source.

10.3 Local reflection models

Very often, computer graphics systems employ simple reflection models to describe how light is reflected by a surface. Local reflection models enable us to calculate the reflected light intensity at a point on a surface, and although it is not correct, it is convenient to divide the light reflected by a surface into three components: ambient, diffuse, and specular. The ambient component represents a constant level of light that is supposed to permeate the illuminated environment. This is included to ensure that surfaces not directly illuminated by virtual light sources receive some light. The diffuse component represents the light reflected by matte surfaces, while the specular component represents the light reflected by shiny surfaces. Obviously, this division of surfaces into three categories is erroneous — nevertheless its simplicity permits reasonably good images to be generated.

Another assumption is that the proportion of light reflected for each light component is represented by three constants:

k_a = the ambient component,
k_d = the diffuse component, and
k_s = the specular component,

where

$$k_a + k_d + k_s = 1.$$

The global ambient light level is defined by I_a, whereas the incident light due to a light source is defined by I_i and gives rise to a diffuse and specular component.

We can now start to derive an equation that sums these three components together. The ambient term is $k_a I_a$ and the diffuse term is $k_d I_i \mathbf{n} \cdot \mathbf{L}$. All that we need now is to compute the specular term, which is viewer-dependent as it simulates the reflection of the light source observed in shiny surfaces.

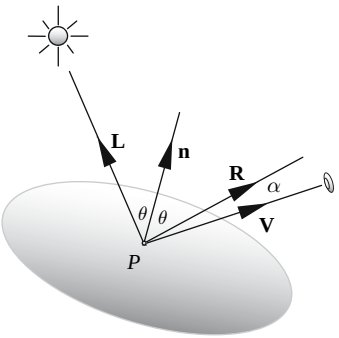


FIGURE 10.3.

Figure 10.3 shows the geometry used to compute the specular component. At a point P on the surface, photons are arriving from the light source L with direction $-\mathbf{L}$. As $\mathbf{n}$ is the surface normal at P, for a perfect reflector the angle of reflection θ will equal the angle of incidence, and the reflected photons will have direction $\mathbf{R}$. If the viewer happens to be looking back along vector $\mathbf{R}$, a bright spot is observed. However, say the viewer is offset by an angle α, and the surface is not a perfect reflector, some light will be seen when looking back along $\mathbf{V}$. Phong [Phong, 1975] proposed that this specular highlight could be controlled by $\cos\alpha$, which gives a bell-shaped distribution of light intensity around $\mathbf{V}$. But if the vectors associated with this model are unit vectors, then

$$\cos\alpha = \mathbf{R} \cdot \mathbf{V}.$$

Phong also suggested that different levels of shininess could be simulated by raising $\cos\alpha$ to some power n:

$$\cos^n \alpha = (\mathbf{R} \cdot \mathbf{V})^n$$

where n is a parameter controlling the level of shininess. For example, when $n = \infty$ (or some very high value), a mirror surface is created, and as n is reduced, the size of the specular highlight increases.

Adding this specular term to the ambient and diffuse terms produces

$$I = k_a I_a + I_i \left(k_d \mathbf{n} \cdot \mathbf{L} + k_s (\mathbf{R} \cdot \mathbf{V})^n\right). \tag{10.1}$$

However, Eq. (10.1) defines intensity, rather than color. The nature of the human eye allows us to describe color in terms of three additive components: red, green blue, which gives rise to three color intensities:

$$I_r = k_{ar} I_{ar} + I_{ir} \left(k_{dr} \mathbf{n} \cdot \mathbf{L} + k_{sr} (\mathbf{R} \cdot \mathbf{V})^n\right)$$
$$I_g = k_{ag} I_{ag} + I_{ig} \left(k_{dg} \mathbf{n} \cdot \mathbf{L} + k_{sg} (\mathbf{R} \cdot \mathbf{V})^n\right)$$
$$I_b = k_{ab} I_{ab} + I_{ib} \left(k_{db} \mathbf{n} \cdot \mathbf{L} + k_{sb} (\mathbf{R} \cdot \mathbf{V})^n\right).$$

(The extra suffices, r, g, and b, identify the red, green, and blue components, respectively.) Having independent terms for each color component enables colored light sources to illuminate different colored surfaces.

Although Eq. (10.1) is correct, it does reference the reflected light vector **R**, which is currently undefined. However, Eq. (5.6) describes how a vector is reflected away from a line or plane:

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2\left(\frac{\mathbf{n} \cdot \mathbf{v}_{\text{in}}}{\|\mathbf{n}\|^2}\right)\mathbf{n}$$

where

$\mathbf{v}_{\text{in}}$ is the incident vector,
$\mathbf{v}_{\text{out}}$ is the reflected vector,
n is the normal vector.

If **n** is a unit vector, this simplifies to

$$\mathbf{v}_{\text{out}} = \mathbf{v}_{\text{in}} - 2(\mathbf{n} \cdot \mathbf{v}_{\text{in}})\mathbf{n}. \tag{10.2}$$

If Eq. (10.2) is to be used in Eq. (10.1) we have to reverse $\mathbf{v}_{in}$ so that it points towards the light source:

$$\mathbf{v}_{\text{out}} = 2(\mathbf{n} \cdot \mathbf{v}_{\text{in}})\mathbf{n} - \mathbf{v}_{\text{in}}$$

and if we substitute the vector names associated with our lighting model, we obtain

$$\mathbf{R} = 2(\mathbf{n} \cdot \mathbf{L})\mathbf{n} - \mathbf{L}. \tag{10.3}$$

Blinn [Blinn, 1977] proposed a way of avoiding calculating **R** altogether. This involved defining a vector **h** half-way between **L** and **V**:

$$\mathbf{h} = \frac{\mathbf{L} + \mathbf{V}}{\|\mathbf{L} + \mathbf{V}\|}. \tag{10.4}$$

And as **h** is viewer–dependent, $\mathbf{n} \cdot \mathbf{h}$ can replace $\mathbf{R} \cdot \mathbf{V}$ in Eq. (10.1):

$$I = k_a I_a + I_i \left(k_d \mathbf{n} \cdot \mathbf{L} + k_s (\mathbf{n} \cdot \mathbf{h})^n\right). \tag{10.5}$$

In fact, Eq. (10.5) is employed in OpenGL and DirectX.

10.4 Shading

The shading models proposed by Gouraud [Gouraud, 1971] and Phong have survived over 30 years of usage. To begin with, Gouraud proposed that if one could compute the light intensity at two opposite edges of a polygon, the light intensity between the two points could be linearly interpolated. Such a scheme is shown in Fig. 10.4.

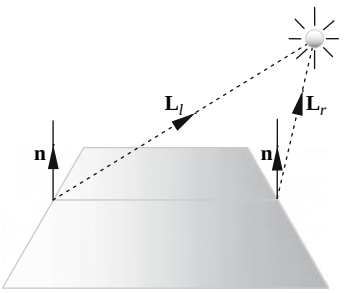


FIGURE 10.4.

On the left-hand side of the polygon the light intensity is calculated as

$$I_l = I_a k_a + I_i k_d \mathbf{n} \cdot \mathbf{L}_l$$

whereas the light intensity on the right-hand side is calculated as

$$I_r = I_a k_a + I_i k_d \mathbf{n} \cdot \mathbf{L}_r.$$

Intermediate points on the line connecting the two extremes are linearly interpolated between I_l and I_r.

Gouraud also devised a clever way of creating the appearance of a smooth object by making the shared edges of polygonal shapes invisible. The strategy involves taking the average normal at each vertex of an object, which ensures that the light intensities along the edges of a polygon are shared by neighbouring polygons. Figure 10.5 shows four connecting polygons with their independent surface normals. An average normal $\mathbf{n}_l$ is computed at the shared vertex where the polygons meet and used in the above lighting calculations:

$$\mathbf{n}_l = \mathbf{n}_1 + \mathbf{n}_2 + \mathbf{n}_3 + \mathbf{n}_4.$$

At the other end of the edge a similar averaging takes place between the four touching polygons to produce an average normal $\mathbf{n}_r$. Light intensities are calculated at either end and then interpolated along the edge. Both normals require normalizing before they are used in the lighting calculations.

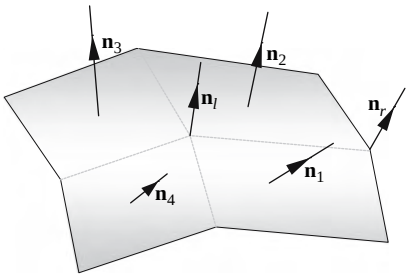


FIGURE 10.5.

Phong's PhD research improved upon Gouraud's diffuse shading model by interpolating the two vectors $\mathbf{n}_l$ and $\mathbf{n}_r$ rather than the light intensities. Although this increases the processing time, the specular highlights achieved across a pseudo-smooth object are worth it. Another part of the extra processing is in normalizing the interpolated normal vector.

10.5 Bump mapping

Irregular surfaces such as bark, or the skin of an orange, have extremely complex surface geometries and present a real challenge to model within a computer graphic system. But as we have seen above, cunning schemes can be invented to achieve an acceptable effect, without expending too much effort, and the same is true for bumpy surfaces.

In 1978 James Blinn published his paper "Simulation of Wrinkled Surfaces" [Blinn, 1978] and invented *bump mapping*, which makes a surface appear irregular without involving extra geometry. Readers are highly recommended to read Mark Kilgard's paper, "A Practical and Robust Bump-Mapping Technique for Today's GPUs" [Kilgard, 2000], upon which the following description is based.

Blinn's technique derives a normal vector at a point on a parametric surface, which is then perturbed using a value from an associated height field (bump map). The surface is then rendered using the perturbed normals, which modulate the surface light intensities accordingly.

We begin by acknowledging some sort of bivariate vector function $\mathbf{P}(u, v)$ that describes a 3D surface, where the parameters (u, v) identify a unique point (x, y, z) on the surface. For example, $\mathbf{P}(u, v)$ might describe a plane, cylinder, sphere, toroid, or parametric patch.

From now on, we will reference this function as $\mathbf{P}$ and remember that its input parameters are (u, v).

By partially differentiating $\mathbf{P}$ with respect to u and v, we obtain two vectors $\partial\mathbf{P}/\partial u$ and $\partial\mathbf{P}/\partial v$, which lie on the tangent plane containing $\mathbf{P}$. By taking their cross product we obtain the surface normal $\mathbf{N}$ at $\mathbf{P}$:

$$\mathbf{N} = \frac{\partial \mathbf{P}}{\partial u} \times \frac{\partial \mathbf{P}}{\partial v}. \tag{10.6}$$

Figure 10.6 illustrates these vectors.

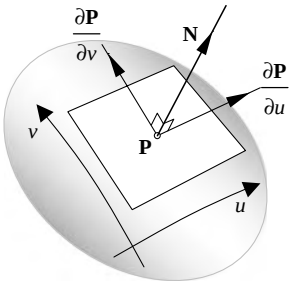


FIGURE 10.6.

The next step is to effectively displace the surface along **N** by a value stored in a bivariate scalar function (bump map) $F(u, v)$. Figure 10.7(a) shows such a map indexed by u and v. Figure 10.7(b) shows a surface defined by a bivariate vector function $\mathbf{P}(u, v)$ with some of its normal vectors. Figure 10.7(c) shows the perturbed surface $\mathbf{P}'(u, v)$ after its normal vectors have been disturbed by $F(u, v)$.

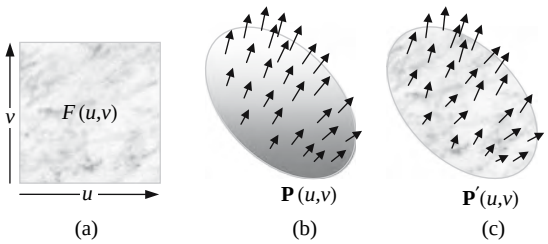


FIGURE 10.7.

Before the displacement is performed, **N** is normalized to keep the process consistent:

$$\frac{\mathbf{N}}{\|\mathbf{N}\|}.$$

Thus, the displaced point $\mathbf{P}'$ is defined as

$$\mathbf{P}' = \mathbf{P} + F\frac{\mathbf{N}}{\|\mathbf{N}\|}.$$

These new points form the secondary perturbed surface that is rendered. But the renderer requires access to the surface normals associated with $\mathbf{P}'$, which is defined using

$$\mathbf{N}' = \frac{\partial \mathbf{P}'}{\partial u} \times \frac{\partial \mathbf{P}'}{\partial v}. \tag{10.7}$$

The partial derivatives in Eq. (10.7) are expanded using the chain rule:

$$\frac{\partial \mathbf{P}'}{\partial u} = \frac{\partial \mathbf{P}}{\partial u} + \frac{\partial F}{\partial u}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right) + F\left(\frac{\partial \frac{\mathbf{N}}{\|\mathbf{N}\|}}{\partial u}\right)$$

and

$$\frac{\partial \mathbf{P}'}{\partial v} = \frac{\partial \mathbf{P}}{\partial v} + \frac{\partial F}{\partial v}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right) + F\left(\frac{\partial \frac{\mathbf{N}}{\|\mathbf{N}\|}}{\partial v}\right).$$

To simplify these partial derivatives, it is assumed that the relative magnitude of F is negligible and can be ignored. Thus,

$$\frac{\partial \mathbf{P}'}{\partial u} = \frac{\partial \mathbf{P}}{\partial u} + \frac{\partial F}{\partial u}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right)$$

and

$$\frac{\partial \mathbf{P}'}{\partial v} = \frac{\partial \mathbf{P}}{\partial v} + \frac{\partial F}{\partial v}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right)$$

which permits Eq. (10.7) to be written as

$$\mathbf{N}' = \left(\frac{\partial \mathbf{P}}{\partial u} + \frac{\partial F}{\partial u}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right)\right) \times \left(\frac{\partial \mathbf{P}}{\partial v} + \frac{\partial F}{\partial v}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|}\right)\right). \tag{10.8}$$

Basically, Eq. (10.8) has the form of four vectors:

$$(\mathbf{a}+\mathbf{b}) \times (\mathbf{c}+\mathbf{d})$$

which, when expanded, equals

$$(\mathbf{a}+\mathbf{b}) \times (\mathbf{c}+\mathbf{d}) = \mathbf{a}\times\mathbf{c} + \mathbf{b}\times\mathbf{c} + \mathbf{a}\times\mathbf{d} + \mathbf{b}\times\mathbf{d}.$$

Expanding Eq. (10.8) along similar lines, we obtain

$$\mathbf{N}' = \frac{\partial \mathbf{P}}{\partial u} \times \frac{\partial \mathbf{P}}{\partial v} + \frac{\partial F}{\partial u}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|} \times \frac{\partial \mathbf{P}}{\partial v}\right) + \frac{\partial F}{\partial v}\left(\frac{\partial \mathbf{P}}{\partial u} \times \frac{\mathbf{N}}{\|\mathbf{N}\|}\right) + \frac{\partial F}{\partial u}\frac{\partial F}{\partial v}\frac{\mathbf{N}}{\|\mathbf{N}\|} \times \frac{\mathbf{N}}{\|\mathbf{N}\|}.$$

But

$$\frac{\mathbf{N}}{\|\mathbf{N}\|} \times \frac{\mathbf{N}}{\|\mathbf{N}\|} = 0$$

and

$$\frac{\partial \mathbf{P}}{\partial u} \times \frac{\partial \mathbf{P}}{\partial v} = \mathbf{N}.$$

Therefore,

$$\mathbf{N}' = \mathbf{N} + \frac{\partial F}{\partial u}\left(\frac{\mathbf{N}}{\|\mathbf{N}\|} \times \frac{\partial \mathbf{P}}{\partial v}\right) + \frac{\partial F}{\partial v}\left(\frac{\partial \mathbf{P}}{\partial u} \times \frac{\mathbf{N}}{\|\mathbf{N}\|}\right)$$

or

$$\mathbf{N}' = \mathbf{N} + \frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right) + \frac{\partial F}{\partial v}\left(\frac{\partial \mathbf{P}}{\partial u} \times \mathbf{N}\right)}{\|\mathbf{N}\|}$$

or

$$\mathbf{N}' = \mathbf{N} + \frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|}. \tag{10.9}$$

Blinn proposed two ways to interpret Eq. (10.9): the first interpretation, and probably the most obvious one, is to see $\mathbf{N}'$ as the sum of two vectors:

$$\mathbf{N}' = \mathbf{N} + \mathbf{D}$$

where

$$\mathbf{D} = \frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|}. \tag{10.10}$$

Figure 10.8(a) shows this vector addition, and Fig. 10.8(b) shows the vector cross product associated with **D**.

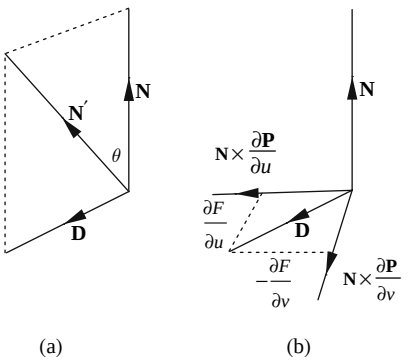


FIGURE 10.8.

Figure 10.8(b) shows us that with **N** as the surface normal vector, $\mathbf{N} \times \partial \mathbf{P}/\partial u$ and $\mathbf{N} \times \partial \mathbf{P}/\partial v$ must lie in the tangent plane of the surface. When these vectors are scaled by the partial derivatives of the bump map F, they are summed to produce **D**, which is subsequently used to perturb **N**.

The height field F was originally used to scale the unit normal vector $\mathbf{N}/\| \mathbf{N} \|$, which is eventually used to perturb **N** by **D**. However, we can save ourselves a lot of work if the bump map takes the form of an offset vector function $\mathbf{D}(u, v)$, then all we have to do is add it to $\mathbf{N}(u, v)$. Such a technique is called an *offset map*.

Blinn's second interpretation of Eq. (10.9) is to imagine that **N** is rotated about some axis to $\mathbf{N}'$. The axis will lie in the tangent plane and is computed using $\mathbf{N} \times \mathbf{N}'$. But as

$$\mathbf{N}' = \mathbf{N} + \mathbf{D}$$

then

$$\mathbf{N} \times \mathbf{N}' = \mathbf{N} \times (\mathbf{N} + \mathbf{D}) = \mathbf{N} \times \mathbf{N} + \mathbf{N} \times \mathbf{D} = \mathbf{N} \times \mathbf{D}. \tag{10.11}$$

If we now substitute Eq. (10.10) in Eq. (10.11), we obtain

$$\mathbf{N} \times \mathbf{N}' = \mathbf{N} \times \mathbf{D} = \mathbf{N} \times \left(\frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|} \right)$$

$$\mathbf{N} \times \mathbf{N}' = \frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right)\right) - \frac{\partial F}{\partial v}\left(\mathbf{N} \times \left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u}\right)\right)}{\| \mathbf{N} \|}. \tag{10.12}$$

Equation (10.12) now contains a triple product, which, as we discovered in Section 2.9.2, has the following identity:

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\,\mathbf{c}.$$

Expanding Eq. (10.12) produces

$$\mathbf{N} \times \mathbf{N}' = \frac{\frac{\partial F}{\partial u}\left(\left(\mathbf{N} \cdot \frac{\partial \mathbf{P}}{\partial v}\right)\mathbf{N} - (\mathbf{N} \cdot \mathbf{N})\frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(\left(\mathbf{N} \cdot \frac{\partial \mathbf{P}}{\partial u}\right)\mathbf{N} - (\mathbf{N} \cdot \mathbf{N})\frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|}.$$

But

$$\mathbf{N} \cdot \mathbf{N} = \| \mathbf{N} \|^2$$

and

$$\mathbf{N} \cdot \frac{\partial \mathbf{P}}{\partial v} = 0 \text{ and } \mathbf{N} \cdot \frac{\partial \mathbf{P}}{\partial u} = 0.$$

Therefore,

$$\mathbf{N} \times \mathbf{N}' = \frac{\frac{\partial F}{\partial u}\left(-\| \mathbf{N} \|^2 \frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(-\| \mathbf{N} \|^2 \frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|} = \|\mathbf{N}\| \left(\frac{\partial F}{\partial v}\frac{\partial \mathbf{P}}{\partial u} - \frac{\partial F}{\partial u}\frac{\partial \mathbf{P}}{\partial v} \right)$$

which can be expressed as

$$\mathbf{N} \times \mathbf{N}' = \| \mathbf{N} \| \mathbf{A} \tag{10.13}$$

where

$$\mathbf{A} = \frac{\partial F}{\partial v}\frac{\partial \mathbf{P}}{\partial u} - \frac{\partial F}{\partial u}\frac{\partial \mathbf{P}}{\partial v}.$$

Because of the cross product in Eq. (10.13), $\mathbf{A}$ must be perpendicular to the plane containing $\mathbf{N}$ and $\mathbf{N}'$. Therefore, one can imagine that $\mathbf{N}$ has been rotated about $\mathbf{A}$ to $\mathbf{N}'$.
Now since

$$\mathbf{N} \times \mathbf{N}' = \mathbf{N} \times \mathbf{D}$$

and

$$\mathbf{N} \times \mathbf{N}' = \| \mathbf{N} \| \mathbf{A}$$

then

$$\mathbf{N} \times \mathbf{D} = \| \mathbf{N} \| \mathbf{A}.$$

Furthermore,

$$\| \mathbf{N} \times \mathbf{D} \| = \| \mathbf{N} \| \| \mathbf{A} \|. \tag{10.14}$$

But as **N** is perpendicular to **D**, we have

$$\| \mathbf{N} \times \mathbf{D} \| = \| \mathbf{N} \| \| \mathbf{D} \| \sin 90^\circ = \| \mathbf{N} \| \| \mathbf{D} \|. \tag{10.15}$$

Equations (10.14) and (10.15) imply that $\| \mathbf{D} \| = \| \mathbf{A} \|$.

N is rotated θ, as shown in Fig. 10.8(a), where

$$\tan\theta = \frac{\| \mathbf{D} \|}{\| \mathbf{N} \|} = \frac{\| \mathbf{A} \|}{\| \mathbf{N} \|}$$

or

$$\theta = \tan^{-1}\left(\frac{\left\| \dfrac{\partial F}{\partial v}\dfrac{\partial \mathbf{P}}{\partial u} - \dfrac{\partial F}{\partial u}\dfrac{\partial \mathbf{P}}{\partial v} \right\|}{\left\| \dfrac{\partial \mathbf{P}}{\partial u} \times \dfrac{\partial \mathbf{P}}{\partial v} \right\|} \right).$$

So this allows us to develop a perturbation strategy based upon rotations, rather than offsets.

Now let's illustrate the above analysis with an example. To keep the mathematics simple, let's use a cylinder as the bivariate vector function, as shown in Fig. 10.9.

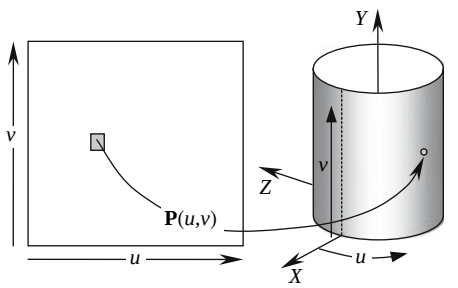


Figure 10.9.

For instance, if the radius and the height of the cylinder equal 1 and $(u, v) \in [0, 2\pi]$, then

the x component is given by $\cos u$,
the y component is given by v,
the z component is given by $-\sin u$,

and

$$\mathbf{P}(u, v) = \cos u\mathbf{i} + v\mathbf{j} - \sin u\mathbf{k}. \tag{10.16}$$

From Eq. (10.16) this makes

$$\frac{\partial \mathbf{P}}{\partial u} = -\sin u\mathbf{i} - \cos u\mathbf{k}$$

$$\frac{\partial \mathbf{P}}{\partial v} = \mathbf{j},$$

$$\mathbf{N} = \frac{\partial \mathbf{P}}{\partial u} \times \frac{\partial \mathbf{P}}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ -\sin u & 0 & -\cos u \\ 0 & 1 & 0 \end{vmatrix}$$

$$\mathbf{N} = \cos u\mathbf{i} - \sin u\mathbf{k}.$$

Figure 10.10 illustrates the orientation of the three vectors $\frac{\partial \mathbf{P}}{\partial u}$, $\frac{\partial \mathbf{P}}{\partial v}$, and $\mathbf{N}$.

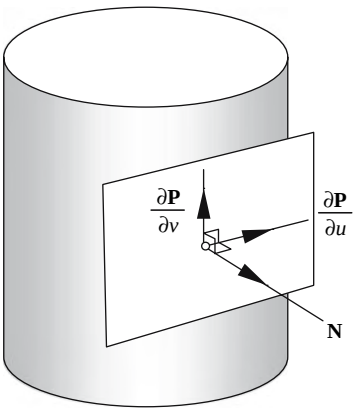


Figure 10.10.

Now let's assume that the height field $F(u, v) \in [0, 255]$ and contains the values shown in Fig. 10.11.

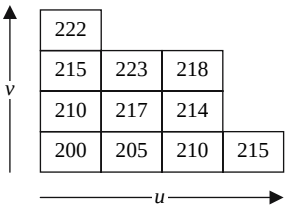


Figure 10.11.

Again, to simply the mathematics, let's find out what happens at the point $u = v = 0$.
To begin with,

$$\frac{\partial \mathbf{P}}{\partial u} = -\mathbf{k}$$
$$\frac{\partial \mathbf{P}}{\partial v} = \mathbf{j}$$
$$\mathbf{N} = \mathbf{i}.$$

From Fig. 10.11, we see that

$$\frac{\partial F}{\partial u} = 205 - 200 = 5$$
$$\frac{\partial F}{\partial v} = 210 - 200 = 10.$$

In calculating $\partial F/\partial u$ and $\partial F/\partial v$, we ignore delta increments in ∂u and ∂v, as this can be compensated for by introducing a scaling factor, which will also control the impact the bump map has on $\mathbf{N}$. Let this scaling factor be λ with an initial value of 1.
Modifying Eq. 10.9 with λ produces

$$\mathbf{N}' = \mathbf{N} + \lambda \left(\frac{\frac{\partial F}{\partial u}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v}\right) - \frac{\partial F}{\partial v}\left(\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u}\right)}{\| \mathbf{N} \|} \right). \tag{10.17}$$

Now let's calculate $\mathbf{N} \times \partial \mathbf{P}/\partial v$ and $\mathbf{N} \times \partial \mathbf{P}/\partial u$:

$$\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = \mathbf{k}$$

$$\mathbf{N} \times \frac{\partial \mathbf{P}}{\partial u} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{vmatrix} = \mathbf{j}.$$

Substituting all the above terms in Eq. (10.17), we obtain

$$\mathbf{N}' = \mathbf{i} + \lambda\,(5\mathbf{k} - 10\mathbf{j})$$

and if $\lambda = 1$,

$$\mathbf{N}' = \mathbf{i} - 10\mathbf{j} + 5\mathbf{k}.$$

If this is seen to be visually unacceptable, λ can be set to some suitable value. In fact, let's make $\lambda = 0.01$ and then compute $\mathbf{N}'$ again:

$$\mathbf{N}' = \mathbf{i} - 0.1\mathbf{j} + 0.05\mathbf{k}.$$

So initially $\mathbf{N} = \mathbf{i}$, and after being perturbed, it is $\mathbf{N}' = \mathbf{i} - 0.1\mathbf{j} + 0.05\mathbf{k}$. Let's compute the angle θ between these vectors using the dot product:

$$\theta = \cos^{-1} \frac{(\mathbf{N} \cdot \mathbf{N}')}{\| \mathbf{N} \| \| \mathbf{N}' \|}$$

$$\theta = \cos^{-1} \left(\frac{(\mathbf{i}) \cdot (\mathbf{i} - 0.1\mathbf{j} + 0.05\mathbf{k})}{1.0125} \right) = \cos^{-1} \left(\frac{1}{1.0125} \right) = 9.01°.$$

As mentioned above, it is possible to convert the original height field into an offset map containing the vectors to perturb $\mathbf{N}$, thus saving considerable repetitive computation.

11 Motion

11.1 Introduction

As vectors have always been associated with dynamics, this text concludes with a short chapter on the motion of objects within an animation context. To begin with, we compute the shortest distance between a stationary observer (which we refer to as "you") and a moving object, followed by that between a moving observer (still "you") and a moving object.

11.2 Close encounters of the first kind

Imagine watching an object fly past you at a speed s_O in a direction $\hat{\mathbf{v}}_O$. The nearest the object approaches you is when a line connecting you to the object is perpendicular to $\hat{\mathbf{v}}_O$, as shown in Fig. 11.1.

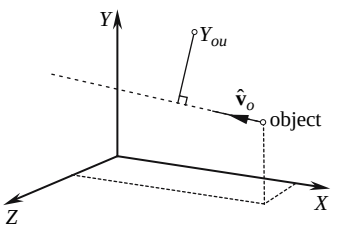


FIGURE 11.1.

If the object is positioned at Q at time $t = 0$, its future position is given by

$$\mathbf{p} = \mathbf{q} + t_P s_O \hat{\mathbf{v}}_O$$

where t_P is the time taken to move from Q to P, as shown in Fig. 11.2.

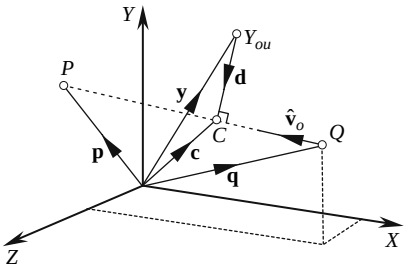


FIGURE 11.2.

The time for the object to reach the "close encounter" at C is t_C, where

$$\mathbf{c} = \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O \tag{11.1}$$

If we let $\mathbf{d} = \overrightarrow{Y_{ou}C}$, then $\mathbf{d} \perp \hat{\mathbf{v}}_O$ and $\mathbf{d} \cdot \hat{\mathbf{v}}_O = 0$.
And as $\mathbf{d} \perp \hat{\mathbf{v}}_O$, $\mathbf{y}$ and $\mathbf{c}$ have identical projections on $\hat{\mathbf{v}}_O$. Therefore,

$$\hat{\mathbf{v}}_O \cdot \mathbf{c} = \hat{\mathbf{v}}_O \cdot \mathbf{y}. \tag{11.2}$$

Multiplying Eq. (11.1) throughout by $\hat{\mathbf{v}}_O$, we obtain

$$\hat{\mathbf{v}}_O \cdot \mathbf{c} = \hat{\mathbf{v}}_O \cdot \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O \cdot \hat{\mathbf{v}}_O. \tag{11.3}$$

Substituting Eq. (11.2) in Eq. (11.3), we get

$$\hat{\mathbf{v}}_O \cdot \mathbf{y} = \hat{\mathbf{v}}_O \cdot \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O \cdot \hat{\mathbf{v}}_O.$$

But $\hat{\mathbf{v}}_O \cdot \hat{\mathbf{v}}_O = 1$. Therefore,

$$\hat{\mathbf{v}}_O \cdot \mathbf{y} = \hat{\mathbf{v}}_O \cdot \mathbf{q} + t_C s_O$$

and

$$t_C = \frac{\hat{\mathbf{v}}_O \cdot (\mathbf{y} - \mathbf{q})}{s_O}. \tag{11.4}$$

The shortest distance between you and the object is given by $\| \mathbf{d} \|$. Now

$$\mathbf{d} = \mathbf{c} - \mathbf{y}.$$

Therefore,

$$\| \mathbf{d} \| = \| \mathbf{c} - \mathbf{y} \|. \tag{11.5}$$

Substituting Eq. (11.1) in Eq. (11.5), we get

$$\| \mathbf{d} \| = \| \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O - \mathbf{y} \|. \tag{11.6}$$

Note that if you are located at the origin, t_C becomes

$$t_C = \frac{-\hat{\mathbf{v}}_O \cdot \mathbf{q}}{s_O} \tag{11.7}$$

and $\| \mathbf{d} \|$ becomes

$$\| \mathbf{d} \| = \| \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O \| \tag{11.8}$$

For example, say you are located at the origin and the object is located at Q, 120 km along the x-axis, as shown in Fig. 11.3.

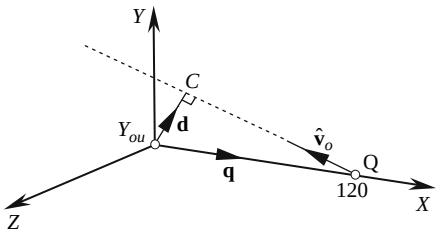


Figure 11.3.

If the object's speed is $s_O = 60\left[\text{km h}^{-1}\right]$ in the direction $\hat{\mathbf{v}}_O = \frac{1}{\sqrt{3}}(-\mathbf{i}+\mathbf{j}+\mathbf{k})$, the time to reach C is given by

$$t_C = \frac{-\frac{1}{\sqrt{3}}(-\mathbf{i}+\mathbf{j}+\mathbf{k}) \cdot (120\mathbf{i})}{60} = \frac{2}{\sqrt{3}} = \frac{2}{3}\sqrt{3}$$

$$t_C = 1.155 \quad [\text{h}].$$

The shortest distance $\| \mathbf{d} \|$ is given by

$$\| \mathbf{d} \| = \| \mathbf{q} + t_C s_O \hat{\mathbf{v}}_O \| = \left\| 120i + \frac{2}{3}\sqrt{3} \times 60 \times \frac{1}{\sqrt{3}}(-\mathbf{i}+\mathbf{j}+\mathbf{k}) \right\|$$

$$\| \mathbf{d} \| = \| 120\mathbf{i} - 40\mathbf{i} + 40\mathbf{j} + 40\mathbf{k} \| = \| 80\mathbf{i} + 40\mathbf{j} + 40\mathbf{k} \|$$

$$\| \mathbf{d} \| = 40 \| 2\mathbf{i} + \mathbf{j} + \mathbf{k} \| = 40\sqrt{6} \cong 98 \quad [\text{km}].$$

11.3 Close encounters of the second kind

The next problem is to consider the scenario when both you and the object are moving! Well, the secret is to translate the problem from one of absolute values to that of relative values. For example, if you are defined as follows:

location: $\mathbf{y}$,
velocity: $s_Y \hat{\mathbf{v}}_Y$,

and the object is defined as

location: $\mathbf{q}$,
velocity: $s_O\hat{\mathbf{v}}_O$,

the object's relationship with you is given by

location: $\mathbf{q}-\mathbf{y}$,
velocity: $s_R\hat{\mathbf{v}}_R = s_O\hat{\mathbf{v}}_O - s_Y\hat{\mathbf{v}}_Y$.

To keep the geometry simple, it is best if you are located at the origin, as shown in Fig. 11.4.

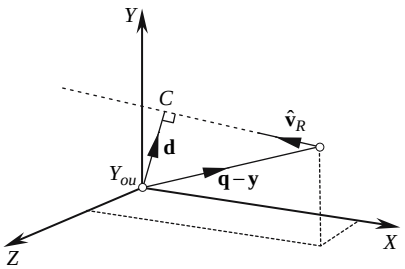


FIGURE 11.4.

With reference to Eq. (11.7), the time taken to reach C is given by

$$t_C = \frac{-\hat{\mathbf{v}}_R \cdot (\mathbf{q}-\mathbf{y})}{s_R} \tag{11.9}$$

With reference to Eq. (11.8), the distance $\| \mathbf{d} \|$ is given by

$$\| \mathbf{d} \| = \| \mathbf{q} + t_C s_R \hat{\mathbf{v}}_R \| \tag{11.10}$$

Let's test Eqs. (11.9) and (11.10) with an example.
Say you are defined as

location: $\mathbf{y} = 10\mathbf{i} + 10\mathbf{j}$ [km],
velocity: $s_Y\hat{\mathbf{v}}_Y = 4\mathbf{i}\ \left[\text{km.h}^{-1}\right]$,

and the object is defined as

location $\mathbf{q} = 25\mathbf{i} + 10\mathbf{j}$ [km],
velocity $s_O\hat{\mathbf{v}}_O = 3\mathbf{j}\ \left[\text{m.h}^{-1}\right]$.

This scenario is shown in Fig. 11.5.

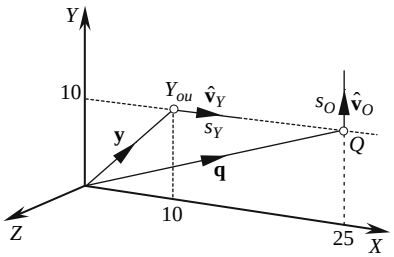


FIGURE 11.5.

The relative values are

relative location of object:

$$\mathbf{q} - \mathbf{y} = 15\mathbf{i}\ [\mathrm{km}]$$

relative velocity of object:

$$s_O\hat{\mathbf{v}}_O - s_Y\hat{\mathbf{v}}_Y = -4\mathbf{i} + 3\mathbf{j}\ \left[\mathrm{km.h}^{-1}\right].$$

Therefore, from the relative velocity, the relative speed is

$$\| s_O\hat{\mathbf{v}}_O - s_Y\hat{\mathbf{v}}_Y \| = \sqrt{(-4)^2 + 3^2} = 5\ \left[\mathrm{km.h}^{-1}\right]$$

and the relative direction is

$$\hat{\mathbf{v}}_O - \hat{\mathbf{v}}_Y = -\frac{4}{5}\mathbf{i} + \frac{3}{5}\mathbf{j}\ [\mathrm{km}].$$

Therefore,

$$t_C = \frac{-\left(-\frac{4}{5}\mathbf{i} + \frac{3}{5}\mathbf{j}\right)\cdot(15\mathbf{i})}{5} = \frac{12}{5} = 2.4\ \left[\mathrm{km.h}^{-1}\right].$$

Substituting 2.4 h in Eq. (11.10) gives

$$\| \mathbf{d} \| = \left\| 15\mathbf{i} + \frac{12}{5} \times 5 \times \left(-\frac{4}{5}\mathbf{i} + \frac{3}{5}\mathbf{j}\right) \right\|$$

$$\| \mathbf{d} \| = \left\| 15\mathbf{i} - \frac{48}{5}\mathbf{i} + \frac{36}{5}\mathbf{j} \right\|$$

$$\| \mathbf{d} \| = \left\| \frac{27}{5}\mathbf{i} + \frac{36}{5}\mathbf{j} \right\| = 9\ [\mathrm{km}].$$

Thus, the close encounter occurs 2.4 hours after the reference time, and the object is 9 km from you.

Appendix A

A.1 Definitions and laws of vector algebra

These are the definitions and laws associated with vectors.

If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors, and a and b are scalars, then

By definition, two vectors $\mathbf{u}$ and $\mathbf{v}$ are equal if they have the same magnitude and direction.

By definition, if two vectors $\mathbf{u}$ and $\mathbf{v}$ have the same magnitude but opposite directions, then $\mathbf{u} = -\mathbf{v}$ or $\mathbf{v} = -\mathbf{u}$.

A.2 Common laws of vector algebra

Commutative law of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
Associative law of addition: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
Commutative law of multiplication: $a\mathbf{u} = \mathbf{u}a$
Associative law of multiplication: $a\,(b\mathbf{u}) = (ab)\,\mathbf{u}$
Distributive law: $(a + b)\,\mathbf{u} = a\mathbf{u} + b\mathbf{u}$
Distributive law: $a\,(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$

By definition, the scalar product $\mathbf{u} \cdot \mathbf{v}$ is $\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta$.

θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

By definition, the vector product $\mathbf{u} \times \mathbf{v}$ is

$$\mathbf{u} \times \mathbf{v} = \mathbf{w}$$

where $\|\mathbf{w}\| = \|\mathbf{u}\| \, \|\mathbf{v}\| \sin\theta$.

θ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

Note: The vector product works only in 3D.

Appendix B

B.1 Vector triple product

In this appendix we will prove that $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}$ and $(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{b}\cdot\mathbf{c})\,\mathbf{a}$, which are both known as *vector triple products*.

In Section 2.10.1 we proved that $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ resides in the plane containing $\mathbf{b}$ and $\mathbf{c}$. Therefore, $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ must be linearly related to $\mathbf{b}$ and $\mathbf{c}$. Thus,

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \lambda\mathbf{b} + \delta\mathbf{c} \tag{B.1}$$

where λ and δ (delta) are scalars. Now we have to expand $\mathbf{a}\times(\mathbf{b}\times\mathbf{c})$ and force the algebra into the form of the RHS of Eq. (B.1). At the same time, we must look out for any hidden dot products. Let's begin the expansion: given $\mathbf{a} = x_a\mathbf{i} + y_a\mathbf{j} + z_a\mathbf{k}$, we have

$$\mathbf{b} = x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}$$
$$\mathbf{c} = x_c\mathbf{i} + y_c\mathbf{j} + z_c\mathbf{k}.$$

Then

$$\mathbf{b}\times\mathbf{c} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_b & y_b & z_b \\ x_c & y_c & z_c \end{vmatrix} = \begin{vmatrix} y_b & z_b \\ y_c & z_c \end{vmatrix}\mathbf{i} + \begin{vmatrix} z_b & x_b \\ z_c & x_c \end{vmatrix}\mathbf{j} + \begin{vmatrix} x_b & y_b \\ x_c & y_c \end{vmatrix}\mathbf{k}$$

and

$$\mathbf{b}\times\mathbf{c} = (y_bz_c - y_cz_b)\,\mathbf{i} + (z_bx_c - z_cx_b)\,\mathbf{j} + (x_by_c - x_cy_b)\,\mathbf{k}.$$

Therefore,

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_a & y_a & z_a \\ y_bz_c - y_cz_b & z_bx_c - z_cx_b & x_by_c - x_cy_b \end{vmatrix}.$$

Expanding gives

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = &[y_a(x_by_c - x_cy_b) - z_a(z_bx_c - z_cx_b)]\,\mathbf{i}\\ &+[z_a(y_bz_c - y_cz_b) - x_a(x_by_c - x_cy_b)]\,\mathbf{j}\\ &+[x_a(z_bx_c - z_cx_b) - y_a(y_bz_c - y_cz_b)]\,\mathbf{k}.\end{aligned}$$

Expanding again gives

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = &(x_by_ay_c - x_cy_ay_b - x_cz_az_b + x_bz_az_c)\,\mathbf{i}\\ &+(y_bz_az_c - y_cz_az_b - x_ax_by_c + x_ax_cy_b)\,\mathbf{j}\\ &+(x_ax_cz_b - x_ax_bz_c - y_ay_bz_c + y_ay_cz_b)\,\mathbf{k}.\end{aligned}$$

Now we split the above equation as closely as possible into the form $\lambda\mathbf{b} + \delta\mathbf{c}$ by isolating the $\mathbf{b}$ and $\mathbf{c}$ components:

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = &\,x_b(y_ay_c + z_az_c)\,\mathbf{i} - x_c(y_ay_b + z_az_b)\,\mathbf{i}\\ &+y_b(z_az_c + x_ax_c)\,\mathbf{j} - y_c(x_ax_b + z_az_b)\,\mathbf{j}\\ &+z_b(x_ax_c + y_ay_c)\,\mathbf{k} - z_c(x_ax_b + y_ay_b)\,\mathbf{k}.\end{aligned} \tag{B.2}$$

We are almost there, but something's still missing. Looking carefully at the expressions $(y_ay_c + z_az_c)$, $(z_az_c + x_ax_c)$, and $(x_ax_c + y_ay_c)$ in Eq. (B.2), we notice that they are similar to the dot-product expansion, except that each is missing a term. These missing terms are x_ax_c, y_ay_c, and z_az_c respectively, and can be introduced by adding

$$x_b(x_ax_c)\,\mathbf{i} + y_b(y_ay_c)\,\mathbf{j} + z_b(z_az_c)\,\mathbf{k}$$

to the RHS of Eq. (B.2). Fortunately, $x_b(x_ax_c)\,\mathbf{i} + y_b(y_ay_c)\,\mathbf{j} + z_b(z_az_c)\,\mathbf{k}$ has to be subtracted from the RHS of Eq. (B.2) to convert the remaining terms into dot-product form. Thus, we get

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = &\,x_b(x_ax_c + y_ay_c + z_az_c)\,\mathbf{i} + y_b(x_ax_c + y_ay_c + z_az_c)\,\mathbf{j} + z_b(x_ax_c + y_ay_c + z_az_b)\,\mathbf{k}\\ &-x_c(x_ax_b + y_ay_b + z_az_b)\,\mathbf{i} - y_c(x_ax_b + y_ay_c + z_az_b)\,\mathbf{j} - z_c(x_ax_b + y_ay_b + z_az_b)\,\mathbf{k}\end{aligned}$$

which contains the following dot products:

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = &\,x_b(\mathbf{a}\cdot\mathbf{c})\,\mathbf{i} + y_b(\mathbf{a}\cdot\mathbf{c})\,\mathbf{j} + z_b(\mathbf{a}\cdot\mathbf{c})\,\mathbf{k}\\ &-x_c(\mathbf{a}\cdot\mathbf{b})\,\mathbf{i} - y_c(\mathbf{a}\cdot\mathbf{b})\,\mathbf{j} - z_c(\mathbf{a}\cdot\mathbf{b})\,\mathbf{k}.\end{aligned} \tag{B.3}$$

Simplifying Eq. (B.3) gives

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{a}\cdot\mathbf{c}\,(x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}) - \mathbf{a}\cdot\mathbf{b}\,(x_c\mathbf{i} + y_c\mathbf{j} + z_c\mathbf{k})$$

which can be written as

$$\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\cdot\mathbf{c})\,\mathbf{b} - (\mathbf{a}\cdot\mathbf{b})\,\mathbf{c}. \tag{B.4}$$

Now let's see what $(\mathbf{a}\times\mathbf{b})\times\mathbf{c}$ expands to. With reference to the above argument, we can state that

$$(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=\begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ y_az_b-y_bz_a & z_ax_b-z_bx_a & x_ay_b-x_by_a\\ x_c & y_c & z_c\end{vmatrix}.$$

Expanding gives

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=&\left[z_c\left(z_ax_b-z_bx_a\right)-y_c\left(x_ay_b-x_by_a\right)\right]\mathbf{i}\\&+\left[x_c\left(x_ay_b-x_by_a\right)-z_c\left(y_az_b-y_bz_a\right)\right]\mathbf{j}\\&+\left[y_c\left(y_az_b-y_bz_a\right)-x_c\left(z_ax_b-z_bx_a\right)\right]\mathbf{k}.\end{aligned}$$

Expanding again gives

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=&\left(x_bz_az_c-x_az_bz_c-x_ay_by_c+x_by_ay_c\right)\mathbf{i}\\&+\left(x_ax_cy_b-x_bx_cy_ay_az_bz_c+y_bz_az_c\right)\mathbf{j}\\&+\left(y_ay_cz_b-y_by_cz_a-x_bx_cz_a+x_ax_cz_b\right)\mathbf{k}.\end{aligned}$$

Again we split the above expansion as closely as possible into the form $\lambda\mathbf{a}+\delta\mathbf{b}$ by isolating the $\mathbf{a}$ and $\mathbf{b}$ components:

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=&\,x_b\left(y_ay_c+z_az_c\right)\mathbf{i}-x_a\left(y_by_c+z_bz_c\right)\mathbf{i}\\&+y_b\left(x_ax_c+z_az_c\right)\mathbf{j}-y_a\left(x_bx_c+z_bz_c\right)\mathbf{j}\\&+z_b\left(x_ax_c+y_ay_c\right)\mathbf{k}-z_a\left(x_bx_c+y_by_c\right)\mathbf{k}.\end{aligned}\tag{B.5}$$

Once more, something's missing. Looking carefully at the expressions $(y_ay_c+z_az_c)$, $(x_ax_c+z_az_c)$, and $(x_ax_c+y_ay_c)$ in Eq. (B.5), we notice that they are similar to the dot-product expansion, except each has a missing term. These missing terms are x_ax_c, y_ay_c, and z_az_c, respectively, and can be introduced by adding

$$x_b\left(x_ax_c\right)\mathbf{i}+y_b\left(y_ay_c\right)\mathbf{j}+z_b\left(z_az_c\right)\mathbf{k}$$

to the RHS of Eq. (B.5). Fortunately, $x_b\left(x_ax_c\right)\mathbf{i}+y_b\left(y_ay_c\right)\mathbf{j}+z_b\left(z_az_c\right)\mathbf{k}$ has to be subtracted from the RHS of Eq. (B.5) to convert the remaining terms into dot-product form. Thus, we get

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=&\,x_b\left(x_ax_c+y_ay_c+z_az_c\right)\mathbf{i}+y_b\left(x_ax_c+y_ay_c+z_az_b\right)\mathbf{j}+z_b\left(x_ax_c+y_ay_c+z_az_b\right)\mathbf{k}\\&-x_a\left(x_bx_c+y_by_c+z_bz_c\right)\mathbf{i}-y_a\left(x_bx_c+y_by_c+z_bz_c\right)\mathbf{j}-z_a\left(x_bx_c+y_by_c+z_bz_c\right)\mathbf{k}\end{aligned}$$

which contains dot products:

$$\begin{aligned}(\mathbf{a}\times\mathbf{b})\times\mathbf{c}=&\,x_b\left(\mathbf{a}\cdot\mathbf{c}\right)\mathbf{i}+y_b\left(\mathbf{a}\cdot\mathbf{c}\right)\mathbf{j}+z_b\left(\mathbf{a}\cdot\mathbf{c}\right)\mathbf{k}\\&-x_a\left(\mathbf{b}\cdot\mathbf{c}\right)\mathbf{i}-y_a\left(\mathbf{b}\cdot\mathbf{c}\right)\mathbf{j}-z_a\left(\mathbf{b}\cdot\mathbf{c}\right)\mathbf{k}.\end{aligned}\tag{B.6}$$

Simplifying Eq. (B.6) gives

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\,(x_b\mathbf{i} + y_b\mathbf{j} + z_b\mathbf{k}) - (\mathbf{b} \cdot \mathbf{c})\,(x_a\mathbf{i} - y_a\mathbf{j} - z_a\mathbf{k})$$

which can be written as

$$(\mathbf{a} \times \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \cdot \mathbf{c})\,\mathbf{b} - (\mathbf{b} \cdot \mathbf{c})\,\mathbf{a}. \tag{B.7}$$

References

Akenine-Möller T. et al. *Real-Time Rendering,* 2nd edition, A. K. Peters, London (2002).

Blinn J. Models of light reflection for computer synthesized pictures. *ACM Computer Graphics (Proc. SIGGRAPH '77)*, pp. 192–198 (1977).

Blinn J. Simulation of wrinkled surfaces. *Computer Graphics (Proc. SIGGRAPH '78)*, pp. 286–292. Also in Tutorial. *Computer Graphics: Image Synthesis*, pp. 307–313 (1978).

Conway J. et al. *On Quaternions and Octonions*, A. K. Peters, London (2003).

Crowe M. *A History of Vector Analysis*, Dover Publications, Inc. (1967).

Glassner A. *An Introduction to Ray Tracing*, Academic Press, New York (1989).

Gouraud H. Continuous shading of curved surfaces. *IEEE Trans. on Computers*, **20**(6), 623–628 (1971).

Kilgard M. A practical and robust bump-mapping technique for today's GPUs. *GDC 2000: Advanced OpenGL Game Development* (2000).

Lewis P. et al. *Vector Analysis for Engineers and Scientists*, Addison-Wesley, Wokingham, UK (1989).

Phong B. Illumination for computer generated pictures. *Comm. ACM*, **18**(6), 311–317 (1975).

Shoemake K. Animating rotation sequences with quaternion curves, SIGGRAPH 85, *Proc. Computer Graphics,* **19**(3), 245–254 (1985).

Vince J. *Mathematics for Computer Graphics*, 2nd edition. Springer, New York (2006).

Further Reading

N. Dodgson, Ray-tracing primitives, http://www.cl.cam.ac.uk/Teaching/2000/AGraphHCI/SMEG/node2.html.

J. Kuipers, *Quaternions and Rotation Sequences*, Princeton University Press, Princeton, NJ (2002).

M. Spiegel,*Theory and Problems of Vector Analysis*, McGraw-Hill, New York (1959).

J. Vince, *Geometry for Computer Graphics,* Springer, New York (2005).

Index

addition
  quaternions, 188
  vectors, 4
algebra
  complex numbers, 183
  vectors, 247
angle/ angles
  between a line and a plane, 114
  between planes, 113
  between vectors, 25
associative law
  vectors, 247
axial system
  change of, 54

back-facing polygon, 42
barycentric coordinates, 160
Bézier quadratic patch, 207
bump mapping, 231

column vector, 12
commutative law
  vectors, 34
complex numbers, 182
  as rotators, 187
  conjugate, 184
  graphical interpretation, 185
  magnitude, 184
cosine/ cosines, 186
Cramer's rule, 162, 176
cross product, vectors, 27

direction cosines, 52
distributive law
  vectors, 34
dot product, 22

ellipse, 138
ellipsoid, 149
equation/ equations
  circle, 133
  ellipse, 138
  ellipsoid, 149
exponential, 186

gimbal lock, 179

Hamilton, 28, 188

inside
  a polygon, 69
  a triangle, 166
  a volume, 112
interpolation, 48
  linear, 48
  quaternions, 49
  spherical, 49
  vectors, 48
intersection/ intersections, 129
  line and a circle, 133
  line and a cone, 158
  line and a cylinder, 152
  line and an ellipse, 138

line and an ellipsoid, 149
line and a plane, 144
line and a sphere, 146
line and a triangle, 160
line segments, 97
lines, 130, 142
planes, 175

light sources, 225
lighting calculations, 225
line/ lines
equation Cartesian form, 65
equation general form, 68
equation parametric form, 61
equidistant from two points, 93
intersecting, 94, 130
intersecting a circle, 133
intersecting a cone, 158
intersecting a cylinder, 152
intersecting a plane, 144
intersecting a sphere, 146
intersecting a triangle, 160
intersecting an ellipse, 138
intersecting an ellipsoid, 149
perpendicular to a vector, 74
reflecting of a line, 123
reflecting of a plane, 126
segment, 94
segments intersecting, 97
skew, 140
linear interpolation, 48
local reflection models, 227

magnitude, vector, 13
Möbius, 160

normal vector, 33

offset map, 234

perp, 44
perpendicular vectors, 44
perspective projection, 213
plane, 101
between two points, 120
perpendicular to a line, 106
through two points, 108

plane equations
Cartesian, 101
from three points, 104
parametric form, 103
plane/ planes, 175
angle between, 113
intersecting, 175
touching a sphere, 171
point/ points
reflected in a line, 86
point on a line
nearest to a point, 81
position vector, 19
product/ products
scalar, 22
vector, 27
projection, 213
fish-eye, 223
perspective, 213

quadratic Bézier patch, 207
quaternions, 188
addition, 188
as rotators, 190
complex conjugate, 192
definition, 189
Hamilton's rules, 188
interpolation, 49
inverse, 195
magnitude, 193
matrix, 197
multiplication, 189
norm, 193
rotating a vector, 195

reflection/ reflections, 86, 123
line off a line, 123
line off a plane, 126
point in a plane, 118
rendering, 225
roll, 221
rotating a vector, 179

scalar product, 22
scalars, 1
shading, 230
Gouraud, 230
Phong, 231
shortest distance, 140

sine, 186
  rule, 49
space
  partitioning 2D, 69
  partitioning 3D, 110
sphere intersecting a plane, 167
sphere touching a plane, 171
sphere touching a triangle, 172
spherical interpolation, 49
straight lines, 61
subtraction, vectors, 4
surface normals, 33

touching, 172
transformations
  change of axes, 55
  perspective, 213
  pitch, 221
  roll, 221
  yaw, 221
triple products, 36, 41, 249

unit vector, 17

vector/ vectors, 11
  addition, 4
  algebra, 14, 34
  Cartesian, 11
  column, 12
  derivative, 201
  differentiation, 201
  direction, 3
  dot product, 22
  interpolating, 48
  length, 13
  magnitude, 13
  non-collinear, 7
  normal, 204
  notation, 2
  perpendicular, 44
  position, 19
  product, 22
  rectangular, 18
  representation, 11
  rotating about an axis, 179
  scalar product, 22
  scalar triple product, 36
  scaling, 6
  spherical interpolation, 49
  subtraction, 4
  unit, 17
  vector product, 27
  vector triple product, 36, 41, 249
volume
  parallelpiped, 37
  tetrahedron, 42